DOCUMENT RESUME

ED 361 402 TM 020 505

AUTHOR Graesser, Arthur C.
TITLE Questioning Mechanisms during Tutoring, Conversation, and Human-Computer Interaction.
INSTITUTION Memphis State Univ., TN. Office of Sponsored Programs.
SPONS AGENCY Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.
PUB DATE Jun 93
CONTRACT N00014-92-J-1826
NOTE 48p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Computer Assisted Instruction; Dialogs (Language); Educational Cooperation; Feedback; Grade 7; *Graduate Students; Higher Education; High Schools; *High School Students; Junior High Schools; Junior High School Students; *Questioning Techniques; Remedial Instruction; Teacher Student Relationship; *Teaching Methods; *Tutoring; *Tutors; Undergraduate Students

ABSTRACT

One-to-one tutoring is more effective than alternative training methods, yet there have been few attempts to examine the process of naturalistic tutoring. Dialogue patterns in tutoring were explored for the following 2 populations: 3 graduate students tutoring 25 college students in research methods, and 10 high school students tutoring 13 seventh graders. Pedagogical strategies, feedback mechanisms, question asking, question answering, and pragmatic assumptions were studied during the tutoring process. One pervasive dialogue pattern was a five-step frame: (1) tutor asks a question; (2) student answers a question; (3) tutor gives short feedback on answer quality; (4) tutor and student collaboratively improve answer quality; and (5) tutor assesses student understanding of the answer. Tutor questions were primarily motivated by curriculum, scripts, and the process of coaching students through exemplar problems. They were rarely motivated by attempts to diagnose and remediate the student's idiosyncratic knowledge deficits. Dialogue patterns were simulated by two computational models--a recurrent connectionist network, and a recursive transition network. These models capture the systematicity in the sequential ordering of speech act categories. Eight tables and three figures present study findings. Attachments list publications of the principal investigator and co-investigators during the grant period. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED 361 402

QUESTIONING MECHANISMS DURING TUTORING, CONVERSATION, AND HUMAN-COMPUTER INTERACTION

Arthur C. Graesser, Principal Investigator

Department of Psychology, Department of Mathematical Sciences, and the Institute for Intelligent Systems

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Mailing address:

Arthur C. Graesser
Department of Psychology
Memphis State University
Memphis, TN 38152
(901) 678-2742
fax: 901-678-2579
graesserac@memst.memstvx1.edu

Reproduction in whole or part is permitted for any purpose of the United States Government.

This research was sponsored by the Cognitive Science Program of the Office of Naval Research, under Contract Number N00014-92-J-1826.

APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED

TM020505

BEST COPY AVAILABLE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1993, June	FINAL REPORT: 15 June 92 to 14 June 93

4. TITLE AND SUBTITLE
QUESTIONING MECHANISMS DURING TUTORING, CONVERSATION, AND HUMAN-COMPUTER INTERACTION

5. FUNDING NUMBERS
N00014-92-J-1826

6. AUTHOR(S)
ARTHUR C. GRAESSER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
OFFICE OF SPONSORED PROGRAMS
MEMPHIS STATE UNIVERSITY
MEMPHIS, TN 38152

8. PERFORMING ORGANIZATION REPORT NUMBER
ONR TECH REPORT
93-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
COGNITIVE SCIENCE PROGRAM (1142 CS)
OFFICE OF NAVAL RESEARCH
800 N. QUINCY STREET
ARLINGTON, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
R&T 4422576

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

One-to-one tutoring is more effective than alternative training methods, yet there have been few attempts to examine the process of naturalistic tutoring. This project explored dialogue patterns in two corpora: graduate students tutoring undergraduates in research methods, and high school students tutoring 7th graders in algebra. We analyzed pedagogical strategies, feedback mechanisms, question asking, question answering, and pragmatic assumptions during the tutoring process. One pervasive dialogue pattern was a five-step frame: (1) tutor asks question, (2) student answers question, (3) tutor gives short feedback on answer quality, (4) tutor and student collaboratively improve on answer quality, and (5) tutor assesses the student's understanding of the answer. Tutor questions were primarily motivated by curriculum scripts and the process of coaching students through exemplar problems -- rarely by attempts to diagnose and remediate the student's idiosyncratic knowledge deficits.

Dialogue patterns were simulated by two computational models: a recurrent connectionist network and a recursive transition network. These models capture the systematicity in the sequential ordering of speech act categories. That is, to what extent does a model accurately predict the category of speech act N+1, given speech acts 1 through N?

14. SUBJECT TERMS	15. NUMBER OF PAGES
QUESTION ASKING, QUESTION ANSWERING, TUTORING, LEARNING	52
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18
298-102

ABSTRACT

One-to-one tutoring is more effective than alternative training methods, yet there have been few attempts to examine the process of naturalistic tutoring. This project explored dialogue patterns in two corpora: graduate students tutoring undergraduates in research methods, and high school students tutoring 7th graders in algebra. We analyzed pedagogical strategies, feedback mechanisms, question asking, question answering, and pragmatic assumptions during the tutoring process. One pervasive dialogue pattern was a five-step frame: (1) tutor asks question, (2) student answers question, (3) tutor gives short feedback on answer quality, (4) tutor and student collaboratively improve on answer quality, and (5) tutor assesses the student's understanding of the answer. Tutor questions were primarily motivated by curriculum scripts and the process of coaching students through exemplar problems -- rarely by attempts to diagnose and remediate the student's idiosyncratic knowledge deficits.

Dialogue patterns were simulated by two computational models: a recurrent connectionist network and a recursive transition network. These models capture the systematicity in the sequential ordering of speech act categories. That is, to what extent does a model accurately predict the category of speech act N+1, given speech acts 1 through N?

It is well documented that one-to-one tutoring is a better method of training students than normal pedagogical strategies in classroom settings. The effect size of the advantage of tutoring over classrooms has ranged from .4 to 2.3 standard deviation units (Bloom, 1984; Cohen, Kulik, & Kulik, 1982; Mohan, 1972). However, it is difficult to determine the cause of this advantage until there is a better understanding of the tutoring process.

Unfortunately, only a handful of studies have systematically examined the process of tutoring at a fine-grained level (Fox, 1992; Graesser, 1992, 1993; Graesser & Person, in press; Leinhardt, 1987; McArthur, Stasz, & Zmuidzinas, 1990; Miyake & Norman, 1979; Putnam, 1987; van Lehn, 1990). It takes a great deal of time and effort to perform an in-depth qualitative analysis of tutorial interaction. Consequently, some of the observations and results reported by these researchers may have limited generality. Because of limited sample sizes in qualitative process-oriented studies, there have been few attempts to relate components of the tutorial process to student achievement or to tutoring outcomes. In the present project, we analyzed patterns of tutorial dialogue in a comparatively large sample of tutoring sessions.

According to Cohen et al.'s (1982) meta-analysis of 52 tutoring studies, the impact of tutoring on learning is not significantly related to the amount of tutoring training that the tutors received. It is also not related to age differences between tutor and student. In some studies, the peers of the students do an excellent job serving as tutors for students having problems (Fantuzzo, King, & Heller, 1992; Mohan, 1972; Rogoff, 1990). These outcomes are rather counterintuitive. Most of us would expect that tutoring age and expertise would improve learning outcomes. One explanation of these results is that the training and expertise of tutors is normally minimal in naturalistic tutoring sessions. Most tutors in a school system are peers of the students, slightly older students, paraprofessionals, and adult volunteers rather than highly skilled tutors (Fitz-Gibbon, 1977). Perhaps a tutor needs extensive training on both the topic knowledge and tutoring strategies before tutoring expertise shows appreciable gains in learning outcomes. Nevertheless, the counterintuitive finding does support one conclusion about the relationship between tutoring process and outcome: The reported facilitation of tutoring over classroom settings can be attributed to pervasive dialogue patterns of normal tutors rather than to special pedagogical strategies of highly trained tutors.

Several hypotheses may explain the advantage of one-to-one tutoring over classroom settings. According to an active inquiry hypothesis, students perhaps have more active control over their learning in tutoring sessions and therefore have a better chance of correcting their own idiosyncratic knowledge deficits. Educational researchers have frequently advocated the construction of educational settings that promote active learning (Bransford, Arbitman-Smith, Stein, & Vye, 1985; Brown, 1988; Nathan, Kintsch, & Young, 1992; Papert, 1980; Scardamalia, Bereiter, McLean, Swallow, & Woodruff, 1989; Zimmerman, Bandura, & Martinez-Pons, 1992). Tutoring allegedly supplies such an environment. According to an error-remediation hypothesis, tutoring provides an opportunity for the tutor to diagnose and repair the idiosyncratic misconceptions and knowledge deficits of a particular student (Anderson & Reiser, 1985; Anderson, Conrad, & Corbett, 1989; van Lehn, 1990). Teachers in classrooms have the time to focus on general problems of several students, but rarely the idiosyncratic problems of a particular student. According to an explanatory reasoning hypothesis, tutoring may expose patterns of reasoning and problem solving that a classroom setting cannot furnish because of time and resource limitations. Learning is facilitated to the extent that students construct explanations and justifications of the content in the material to be learned (Anderson et al., 1989; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Cobb, Wood, Yackel, & McNeal, 1992; Keiras, 1992; Moore & Ohlsson, 1992; Pressley, Symons, McDaniel, Snyder, & Turnure, 1988; Reiser, Kimberg, Lovett, & Ranney, 1991). There no doubt are additional hypotheses that account for the advantages of tutoring over classroom settings. The analyses in this project narrowed down the set of plausible hypotheses.

Ideal tutoring strategies have been proposed by researchers investigating the cognitive foundations of complex learning and by developers of intelligent tutoring systems (Bransford, Goldman, & Vye, 1991; Lesgold, 1992; Ohlsson, 1986; Scardamalia et al., 1989; Sleeman & Brown, 1982). These researchers have identified pedagogical techniques that the tutor can implement during tutoring, such as the Socratic method (Collins, 1985), inquiry teaching (Collins, 1988), diagnosis-remediation (Anderson & Reiser, 1985; van Lehn, 1990), the reciprocal training method (Palincsar & Brown, 1984), modeling-scaffolding-

fading (Collins, Brown, & Newman, 1989; Rogoff, 1990), and curriculum scripts (Putnam, 1987). These pedagogical techniques fall somewhere between the extremes of complete student control (i.e., active inquiry by the student) and complete tutor control (i.e., a tutor lecture). However, the extent to which these pedagogical techniques have been used in naturalistic tutoring has yet to be documented. Given that the vast majority of tutors in school systems have received little or no training in tutoring (Fitz-Gibbon, 1977), the sophisticated pedagogical techniques presumably are infrequent.

This ONR project investigated the dialogue patterns in naturalistic tutoring sessions. We analyzed tutorial dialogue as knowledge was collaboratively constructed and modified. In addition to documenting some basic facts about tutorial dialogue, we focused on four components in depth:

1. Question asking and answering. What mechanisms account for the questions and answers of tutors and students?

2. Feedback during the construction of common ground. Does the student give accurate feedback to the tutor on the student's understanding of the material? Does the tutor give the student accurate feedback on the quality of the student's contributions?

3. Dialogue patterns. What are the pervasive dialogue patterns during tutoring? In particular, we will concentrate on a 5-step dialogue frame.

4. Pragmatic assumptions. What pragmatic assumptions are followed during tutoring? To what extent are these assumptions the same as or different from the pragmatic assumptions in everyday conversation?

These aspects of tutorial dialogue may or may not be compatible with the goals of good pedagogy. We will identify ways that tutors might strategically improve learning by changing the normal course of tutorial dialogue.

We reported some analyses of tutoring sessions in previous reports (Graesser, 1992, 1993; Graesser, Person, & Huber, 1992, 1993; Graesser & Person, in press; Person, Graesser, Magliano, & Kreuz, 1993). A final report on our previous ONR grant ("Questioning Mechanisms during Complex Learning", N00014-90-J-1492, R&T 4422548) summarizes earlier analyses of the tutoring data.

Naturalistic Tutoring Sessions: Two Corpora

Research methods corpus

Graduate students in the psychology department at Memphis State University tutored undergraduate students on troublesome topics in a research methods course (offered by the psychology department). All 25 students in the course were tutored as part of a course requirement, so there was a full range of student achievement (i.e., not just underachieving students). The three tutors had received A's in a graduate-level research methods course. Therefore, the corpus involved "cross-age" tutoring, which is one of the common types of tutoring in school systems. The tutors had never tutored in the area of research methods before this study, but they had occasionally tutored on other topics.

There were 44 one-hour tutoring sessions. The tutoring sessions were videotaped and transcribed. The room used for tutoring was equipped with a video camera, a television set, a marker board, colored markers, and the textbook for the course. The camera was positioned so that the student and the entire marker board was in sight. Therefore, the transcripts of the tutoring sessions included both spoken utterances and messages on the marker board. The transcribers were instructed to transcribe the entire tutoring sessions, including all "ums", "ahs", word fragments, broken sentences, and pauses. Messages on the marker board were sketched in as much detail as possible.

The sessions covered six troublesome topics in an undergraduate research methods course. The topics were operational definitions of variables, graphs, inferential statistics, the evolution of hypothesis to

design, factorial designs, and interactions. An index card was prepared for each topic; 3-5 subtopics were listed under each subtopic. The tutor was asked to cover the topic and subtopics on the index card during the course of the tutoring session. The tutors were not given a specific format to follow, but they were told to resist the temptation of simply lecturing to the student. The students were exposed to the material covered on a topic before they participated in a tutoring session. The topic was covered in a classroom lecture by the instructor before the tutoring session. In addition, both the student and the tutor were required to read specific pages in a research methods text before the tutoring session.

Each of the 25 students participated in two tutoring sessions, yielding 50 sessions altogether. Each student was randomly assigned to 2 of the tutors. Six of the 50 sessions could not be analyzed because the voices were not sufficiently audible on the videotape. Thus, analyses were performed on 44 tutoring sessions.

Examination scores and final grades were available for the 25 undergraduate students, so we could investigate the relationship between student achievement and tutoring processes. A total examination score was based on three objective examinations throughout the semester; there was a total of 150 four-alternative forced-choice questions. The 25 students had a mean score of 100.6 (SD = 11.4). Regarding the final grade received in the course, 4 students received an A, 9 received a B, 10 received a C, and 4 received a C- or D.

Algebra corpus

This corpus consisted of 22 tutoring sessions in which high school students tutored 7th graders on troublesome topics in algebra. There were 13 students who were having trouble with particular topics in their algebra course (according to their teachers). There were 10 tutors who normally provided the tutoring services for the middle school. On the average, a tutor had 9 hours of prior tutoring experience before tutoring a student in this sample. The corpus of tutoring sessions included almost all of the tutoring sessions that occurred in the middle school for 7th graders learning algebra during a one month period. Unlike the research methods corpus, the tutoring sessions in this algebra corpus were remedial activities for underachieving students. Unfortunately, grades and test scores were not available for these students, so it was not possible to assess the relationship between achievement and tutoring processes.

Almost all of the tutoring sessions covered three tutoring topics that are frequently problematic to 7th graders. These include (a) calculation of positive and negative numbers, (b) constructing equations from algebra word problems, and (c) fractions. An examination and chapter excerpt from a textbook were normally associated with each topic. The tutoring sessions lasted approximately 60 minutes, which was comparable to the research methods corpus. A research assistant from Memphis State University videotaped the sessions in a similar manner as the sessions were videotaped in the research methods corpus.

Reliability of scoring tutoring transcripts on content variables

Previous reports and articles have discussed how the transcripts were analyzed on content categories (Graesser, 1992; Graesser & Person, in press; Graesser, Person, & Huber, 1992, 1993). Therefore, these details will not be covered in this report. Trained research assistants were capable of reliably coding most of the data: segmenting transcripts into speech act units, assigning speech acts to speech act categories, identifying questions, assigning questions to question categories, identifying mechanisms that generate questions, and classifying tutor feedback. Whenever these categories were scored, two judges independently furnished the judgments and achieved sufficient interjudge reliability (i.e., Cronbach's alpha = .70 or higher).

The judges needed to have more expertise in the case of some coding analyses. One such analysis consisted of the quality of a contribution in a tutoring session. There were four levels of answer quality: (1) error-ridden answer, (2) vague answer or no information, (3) partially correct answer, and (4) completely correct answer. The judges needed to have a high amount of domain knowledge about research methods to make these judgments. Therefore these judgments were made by professors, postdocs, or 4th-year graduate students in experimental psychology. Other analyses that required special

expertise involved global levels of the tutorial dialogue (e.g., whether an excerpt involved the application of a curriculum script, error-remediation, or some other global process). In this case, the judges needed to have sophisticated knowledge about the tutoring process in addition to extensive domain knowledge. A pair of judges collaboratively supplied judgments in the case of dimensions or categories that required high expertise.

Student Contributions in Tutorial Dialogue

Tutorial dialogue is presumably guided or constrained by the knowledge deficits and misconceptions of a particular student. To what extent does the student actively guide tutorial dialogue? Does the tutor accurately infer the level of knowledge and the misconceptions of the student? Is the student capable of detecting his or her own knowledge deficits and level of understanding? This section addresses the role of the student in tutorial dialogue. We present a number of claims, with empirical data backing each claim.

Claim 1: Students rarely control tutorial dialogue.

Students rarely initiate exchanges that exert control over the tutorial dialogue. In the research methods corpus, only 5% of the subtopics were initiated by the student whereas 95% were initiated by the tutor. The corresponding percentages in the algebra sample were 10% and 90%, respectively. When students did initiate a new subtopic, they normally brought up an example problem or concept that they were having difficulty with (e.g., "I had trouble with problem 4", "I don't understand what an antagonistic interaction is"). Students never set the agenda for the tutoring session. In both tutoring corpora, the tutor carried the burden of setting the agenda, introducing subtopics, and proposing problems to solve.

This result is incompatible with the active inquiry hypothesis that was briefly discussed earlier. That is, the advantage of tutoring over classroom settings cannot be attributed to the student taking active control of the learning experience. With rare exceptions, students were not inquisitive, active, self-regulators of their knowledge in these tutoring sessions. Tutors need to impose special strategies of transferring control to the student if there is a commitment to promote active learning. Such strategies were not in the repertoire of the normal tutor.

There was one finding that indicated that students are somewhat more active in tutoring contexts than in classroom settings. Student questions were more frequent in the tutoring settings than in classroom settings (Graesser & Person, in press). The mean number of student questions per hour was 21.1 (SD = 13.0) in the research methods corpus and 32.2 (19.7) in the algebra corpus. In contrast, a particular student in a classroom setting asks only .11 question per hour; an entire class of students asks only 3.0 questions per hour (Dillon, 1988; Graesser & Person, in press). From the standpoint of a single student, student questions were approximately 250 times as frequent in tutoring sessions as in classrooms. In spite of the high incidence of student questions during tutoring, tutor questions were substantially more prevalent than student questions in tutoring sessions. We found that 80% of the questions in a session were asked by the tutor (82% in the research methods corpus and 78% in the algebra corpus). This percentage is somewhat lower than the percentage of teacher questions in a classroom (96%). In summary, student questions are much more prevalent in tutoring sessions than in classrooms, but it is still the tutor who asks most of the questions and thereby governs the course of the session.

Most of the questions that students asked during the tutoring session did not address their own knowledge deficits. Knowledge deficit questions occur under the following conditions: (a) when the student encounters an obstacle in a plan or problem, (b) when the student detects a contradiction, (c) when an unusual or anomalous event is detected, (d) when there is an obvious gap in the student's knowledge base, and (e) when the student needs to make a decision among a set of alternatives that are equally likely (Graesser & McMahen, 1993; Graesser, Person, & Huber, 1992, 1993). Only 29% of the student questions were knowledge-deficit questions (Graesser & Person, in press), which amounts to 7.7 questions per hour. Most of the student questions (54%) were attempts to confirm the validity of their own beliefs (e.g., "Doesn't a factorial design have two independent variables?") or to confirm common ground (e.g., "Are you talking about the second condition?").

Good students did not ask more questions. Good students also did not tend to ask more knowledge-deficit questions. The frequency of student questions was not robustly related to achievement in the research methods corpus. The correlations were low between examination scores and (a) the total number of student questions ($r = -.22$) and (b) the proportion of student questions that addressed knowledge deficits ($r = .15$). The correlations were also low when final grade was the measure of achievement ($r = -.34$, $p < .05$ for total number of questions; $r = .32$ for proportion of questions that involved knowledge deficits). Other researchers have also failed to show a positive relationship between question asking and achievement (Fishbein, Eckart, Lauver, van Leeuwen, & Langmeyer, 1990).

In summary, the available evidence supports claim 1. Students rarely take an active role in governing the agenda in the tutoring session. They rarely expose their own knowledge deficits and actively seek remediation. Students ask far fewer questions than tutors and most of their questions do not address their knowledge deficits. It is not the case that the good students are more active and ask more questions. Students apparently need to be trained how to ask questions and to be active learners. It is the tutor who carries the burden of establishing the tutoring agenda, introducing topics, presenting examples to work on, and exposing the student's knowledge deficits. The active inquiry hypothesis does not explain why learning is better in one-to-one tutoring than classroom settings.

Claim 2: Deep reasoning questions are prevalent in tutoring sessions.

There is extensive evidence that comprehension improves if students are trained how to ask good questions and to seek answers to the questions (King, 1989, 1992; Rosenshine & Chapman, 1990; Singer & Donlan, 1982; Wong, 1985). However, the process of asking good questions does not come naturally to students, so they need to be trained in developing this cognitive skill (Pressley, 1990). Therefore, we investigated the quality of questions in the tutoring protocols.

One index of question quality is whether the question exposes deep reasoning about the problems and domain topics. In logical reasoning, the statements expressed in an answer consist of the premises and conclusions of a logical syllogism. In causal reasoning, the answer conveys the antecedents and consequences of events. In goal-oriented reasoning, the answer traces the goals and planning of agents. It is well documented that comprehension and memory for technical material improves to the extent that the learner constructs explanations and justifications (Chi et al., 1989; Cobb et al., 1992; Pressley et al., 1988). According to the explanatory reasoning hypothesis discussed earlier, tutoring facilitates learning because it exposes explanations and justifications.

Graesser's question taxonomy specifies those question categories that expose deep reasoning (Graesser & Person, in press; Graesser, Person, & Huber, 1992, 1993). They include the following six categories.

1. Antecedent questions (why?, how?). What caused a state or event? What logically explains or justifies a proposition?

2. Consequence questions (what if?, what next?). What are the causal consequences of a state or event? What are the logical consequences of a proposition?

3. Goal orientation (why?). What are the goals or motives behind an agent's action?

4. Enablement (why?, how?). What object or resource allows an agent to perform an action? What state or event allows another state or event to occur?

5. Instrumental/procedural? (how?). What instrument or plan allows an agent to accomplish a goal?

6. Expectational (why not?). Why did an expected state or event not occur? Why didn't an agent do something?

These questions are manifested in a tutoring session to the extent that the tutor and student explore deeper levels of comprehension. It should be noted that these deep reasoning questions were highly correlated with the deeper levels of cognition in Bloom's taxonomy of educational objectives in the cognitive domain (Bloom, 1956), $r = .64$, $p < .05$. Low-level questions in Bloom's taxonomy inquire about specific facts, terminology, and explicit information in a text; deeper level questions involve reasoning, application, analysis, synthesis, and evaluation (see also Scardamalia & Bereiter, 1992).

Our analysis of the research methods corpus and algebra corpus uncovered an impressive number of deep reasoning questions. The proportion of student questions that were deep reasoning questions was .22 in the research corpus and .39 in the algebra corpus; the corresponding proportions for tutor questions were .16 and .17, respectively. In a typical tutoring session, a student asked approximately 8 deep reasoning questions (per hour) whereas a tutor asked 19 questions. The incidence of deep reasoning questions was much higher in the tutoring sessions than in normal classroom settings, according to our best estimates from published studies on classroom questioning (Dillon, 1988; Graesser & Person, in press). The incidence of student questions in a classroom is extremely low in all published studies (.11 question per student per hour), so deep reasoning questions would also be low. Only 4% of the teacher questions in a classroom are deep questions in Bloom's taxonomy; the vast majority of teacher questions are short answer questions that grill students on explicit material (Dillon, 1988; Kerry, 1987). Therefore, the explanatory reasoning hypothesis provides a very plausible account of the finding that learning is better in tutoring than in classroom settings.

The good students asked a higher proportion of deep reasoning questions. There was a significant positive correlation between the proportion of student questions that were deep reasoning questions and (a) examination scores ($r = .44$, $p < .05$) and (b) final grades ($r = .58$, $p < .05$). Therefore, good students penetrated the deeper levels of comprehension.

Although the incidence of deep reasoning questions is quite high in tutoring sessions, we believe that the quality of student questions and tutor questions could substantially improve. Most of the students' deep reasoning questions were in the instrumental/procedural category (.59 in the research methods corpus and .74 in the algebra corpus). This is the least sophisticated category of the deep reasoning questions. The student is merely requesting that the tutor describe how to compute a function or perform a procedure (e.g., "How do you solve this problem?"). The student might learn how to apply a formula or procedure mechanically, without any understanding of the reasons, justifications, and principles behind each step (Cobb et al., 1992; Greeno, 1982; Mayer, 1992; Ohlsson & Rees, 1991). Given that one of the contemporary missions of the National Council of Teachers of Mathematics (1989) is to promote learning with understanding, one approach to meeting this objective is to teach better question asking skills.

We have developed computer software that requires students to ask questions and that exposes them to good questions. Our "Point and Query" (P&Q) software forces students to learn entirely by asking questions and reading answers to the questions (Graesser, Langston, & Lang, 1992; Graesser, Langston, & Baggett, 1993). In order to ask a question, the student first points to a word or picture element on the computer screen and then to a question that is relevant to the element (from a menu of relevant questions). The menu of relevant questions is formulated on the basis of background knowledge structures and a theory of human question answering called QUEST (Graesser & Franklin, 1990; Graesser, Gordon, & Brainerd, 1992; Graesser & Hemphill, 1991; Graesser, Lang, & Roberts, 1991). The P&Q system is similar to some other menu-based question asking systems that have been developed (Schank, Ferguson, Birnbaum, & Greising, 1991; Sebrechts & Swartz, 1991). The incidence of student questions is quite high on the P&Q software. Whereas a student asks .1 question per hour in a classroom and 27 questions per hour in a tutoring session, the student asks 135 questions per hour when using the P&Q software.

The P&Q software is a promising environment for teaching question asking skills. The quality of the students' questions should improve by exposing them to good questions on the question menu. After extensive experience with the P&Q software, students would automatize good question asking skills. This might have a radical impact on improving comprehension because, as discussed earlier, there is extensive evidence that comprehension improves after students are trained how to ask good questions.

Claim 3: Students reveal their knowledge in their answers to topic-relevant questions.

Ideally, the tutor should be able to adjust the level of instruction and remediation to the idiosyncratic knowledge deficits and misconceptions of a particular student. This requires the tutor to have a valid way of assessing what the student understands. The developers of many intelligent tutoring systems, for example, have embraced student modeling as an important principle of ITS design (Anderson & Reiser, 1985; Burton & Brown, 1982; Clancey, 1983; Ohlsson, 1986; Van Lehn, 1990). Hence the question arises: How does the tutor accurately infer what the student knows? We performed some analyses on the research methods corpus in order to determine whether the students' achievement is reflected in their questions and their answers to questions.

Table 1 presents correlations between student achievement and several measures of student questions and answers. Consider first the measures that do not correlate with achievement. Tutors did not accurately infer student knowledge on the basis of the frequency of student questions or the proportion of student questions that were knowledge-deficit questions. These correlations were either nonsignificant or marginally significant at a lax alpha-level.

Tutors also could not accurately gauge student understanding by merely asking the students (e.g., "Do you understand?", "Do you follow?", "Okay?"). When these comprehension-gauging questions are asked, the student either answers YES ("I understand"), answers NO ("I don't understand"), or gives an indecisive response (no answer, "I don't know"). Are these answers a valid reflection of the student's true understanding? The data revealed that they are not accurate. There was a near zero correlation between student achievement and the likelihood of the students' answering YES. In fact, this relation was found to be significantly curvilinear, .46, .62, .61, and .52 for students receiving final grades of A, B, C, and C-/D, respectively. This was the only significant curvilinear relationship in all of the correlational analyses involving the measures in Table 1. Regarding the NO answers, there was a significant positive correlation between exam scores and the likelihood of students' answering NO (I don't understand). This is a counterintuitive outcome: It was the good students who tended to say that they did not understand. Chi et al. (1989) also reported a positive correlation in the domain of physics between student understanding and the likelihood of students answering NO. Therefore, available evidence indicates that a tutor cannot simply ask students whether they understand and expect the students to supply accurate feedback. The feedback is misleading. Students are very poor at calibrating their own comprehension of material (Glenberg, Wilkinson, & Epstein, 1982; Weaver, 1990).

According to Table 1, there was a robust correlation between achievement and the proportion of student questions that were deep reasoning questions. This correlation was discussed earlier. We suspect, however, that it would be difficult for the tutor to gauge student understanding by this index. An average student asks only 8 deep reasoning questions per hour, so the tutor would be basing the computation on a low frequency event. Although good students had a higher proportion of deep reasoning questions than poor students, the absolute frequency of deep reasoning questions did not significantly vary with student achievement (because good students tended to ask fewer questions). It would indeed be a very subtle cognitive computation for the tutor to estimate the proportion of student questions that are deep reasoning questions. We conclude that the occurrence of students' deep reasoning questions does not provide a reliable basis for inferring student knowledge.

The students' answers to topic-related questions provided the most reliable basis for inferring student knowledge. There was a robust negative correlation between student achievement and the proportion of students' answer contributions that were in the categories of error-ridden, vague, or no-answer. There was a positive correlation between achievement and student answers that were completely correct. It should be noted that tutors asked a large number of questions (104 questions per hour), so there was ample opportunity for the students to give answers and for the tutor to evaluate the quality of the answers. Therefore, it is the tutor's burden to judiciously select questions that diagnose the student's knowledge deficits, bugs, and deep misconceptions.

Tutor Contributions in Tutorial Dialogue

We have established that the tutor plays the primary role in setting the agenda, introducing topics, selecting exemplar problems, and asking questions. In fact, 90-95% of the new topics and problems were initiated by the tutor in the research methods corpus and the algebra corpus. The tutor asked 78-82% of the questions. The tutor established the ground rules and format in all of the tutoring sessions. This section identifies the pedagogical strategies and dialogue patterns that were implemented by the tutor.

Claim 4: Sophisticated tutoring strategies are rare.

Tutors rarely implemented sophisticated tutoring strategies, such as the Socratic method (Collins, 1985), inquiry teaching (Collins, 1988), the reciprocal training method (Palincsar & Brown, 1984), and modeling-scaffolding-fading (Collins et al., 1989; Rogoff, 1990). These methods were virtually nonexistent in the research methods corpus and the algebra corpus. It takes a large amount of training and experience for tutors to use these sophisticated pedagogical strategies. It is therefore not surprising that the strategies were nonexistent in our sample of 13 tutors, and presumably are nonexistent in real school settings. There should be high payoffs in learning outcomes for those researchers and practitioners who introduce sophisticated tutoring strategies in research projects and in school curricula.

Claim 5: Most of the tutors' activities and questions are generated by curriculum scripts.

We analyzed a sample of tutor questions in order to determine what mechanisms generate tutor questions and what agenda is set by the tutor. We selected 249 questions from the research methods corpus and 93 questions from the algebra corpus. Approximately half of the questions were deep reasoning questions (as defined earlier) and half were short-answer questions (e.g., concept completion, quantification, feature specification). For each of these questions, we identified one or two mechanisms that generated the question (see Table 2). We also specified how the tutorial dialogue continued after the tutor question was answered (see Table 3). The latter analysis provides a snapshot of the typical agenda set by the tutor or initiated by the student.

The data in Tables 2 and 3 support the conclusion that the tutors' curriculum scripts generated most of the tutor questions, new subtopics, and tutoring activities. The curriculum script consists of a set of subtopics, examples, and questions that the tutor selects for the tutoring session (Putnam, 1987). In the case of the research methods corpus, the tutor selected the subtopics in a top-down fashion. The selected subtopics had a close correspondence to the information in the chapter excerpts and the index cards supplied by the experimenter (with the major topic and 3-5 subtopics). Virtually all of the examples selected by the tutor came directly from the book. Very often a tutor introduced the same example, subtopic, or question to several students that were tutored on a particular topic. Most (67%) of the questions were asked in the context of an example problem in the research methods course. Examples played an even more predominant role in the algebra corpus; 92% of the tutor questions were asked in the context of a specific example. The tutor normally selected a problem from the student's examination or textbook. After the tutor selected the example problem, the tutor typically coached the student to a solution, or the tutor and student collaboratively solved the problem. It should be noted that the curriculum script is not necessarily a rigid structure in terms of the selection of material and the ordering of material. According to McArthur et al. (1990), the tutor revises and replans the agenda throughout the course of the tutoring session. The revision and replanning are no doubt influenced by the student's performance.

Claim 6: Very few of the tutors' questions and activities are triggered by student errors and misconceptions.

The results in Tables 2 and 3 support this claim. The tutor did not spend much time diagnosing, dissecting, and troubleshooting the student errors that were manifested in the dialogue. According to diagnosis-remediation models of intelligent tutoring (Anderson & Reiser, 1985; van Lehn, 1990), the tutor should spend time diagnosing and correcting the student's conceptual bugs and misconceptions. These

bugs and misconceptions are manifested by the errors committed by the student. As will be reported later, the tutor does normally correct errors that surface. However, the tutor does not spend much time rectifying the buggy rules and deep misconceptions that explain the errors. It is very difficult for a tutor to identify the underlying bugs and misconceptions, let alone to repair them. Consequently, tutors do not normally invest the time in such activities.

Claim 7: A 5-step dialogue frame is a pervasive dialogue pattern.

An extremely pervasive dialogue pattern consisted of a 5-step dialogue frame that was initiated by a tutor question.

Step 1: Tutor asks question
Step 2: Student answers question
Step 3: Tutor gives short feedback on the answer
Step 4: Tutor improves the quality of the answer by directly supplying information or by initiating a collaborative exchange
Step 5: Tutor assesses the student's understanding of the answer

Figure 1 specifies further the components of this dialogue frame. An example of this frame is provided below:

1. TUTOR: Now what is a factorial design?
2. STUDENT: The design has two variables.
3. TUTOR: Uh-huh.
4. TUTOR: So there are two or more independent variables and one dependent variable.
5. TUTOR: Do you see that?
   STUDENT: Uh-huh.

In step 1, the tutor normally asks a single question. Sometimes the question is not posed clearly or as intended, so the tutor revises the question. Successive tutor questions drift systematically in a manner that makes it easier for the student to answer the question (Graesser, 1992). For example, in the excerpt below, an answer to the first question would involve an elaborate construction of information, whereas a simple YES or NO would be an adequate answer to the second question.

TUTOR: So how could we do that [operationally define intelligence]? I mean, do you think that everyone agrees on what intelligence is?

In the following example, the tutor restates the question in different words that provide a more succinct focus on the intended question. It illustrates that the process of constructing a question is iteratively distributed over time.

TUTOR: Did you see how they did that? How did they manage to do that? What did they do there?

Sometimes the student does not understand the question, particularly when the question is not adequately specified. The student asks a counter-clarification question to gain clarity on what the question is. The tutor answers the embedded counter-clarification question and then the student answers the original question. This is illustrated in the excerpt below.

TUTOR: Why would a researcher even want to use more than two levels of an independent variable in an experiment?
STUDENT: More than two levels?
TUTOR: Uh huh.
STUDENT: They would, um, it'd be real accurate 'cause it would show if there's a curvilinear.

In step 2, the student produces an answer to the question. The process of the student constructing an answer is iteratively constructed over time, as the above example illustrates. Answers are not immediately

articulated in a clear, succinct, coherent form. The student frequently produces single words or incoherent fragments of information. The tutor ends up working with these fragments (in step 4) in a fashion that allows a reasonable answer to evolve. When a student's initial answer is incomplete, the tutor frequently pumps the student for additional information by expressing neutral feedback in step 3 (e.g., "uh huh"). There is an iteration of steps 2 and 3 when the tutor pumps the students for more answer information.

In step 3, the tutor gives short feedback on the student's answer. The feedback is positive, negative, or neutral. Most of the time the feedback is expressed verbally. Occasionally the tutor nods or shakes his head to express feedback. When the feedback is neutral on the written transcript, it is necessary to view the videotape and code the intonation of the utterance in order to accurately classify the feedback as positive, negative, versus neutral (Fox, 1992). We have found that 34% of the neutral observations on the written transcripts ended up being either positive or negative when the videotape was viewed. Tutors rarely used lengthy pauses or hesitations to signify negative feedback. The likelihood of the tutor pausing or hesitating in step 3 did not vary as a function of the quality of the student's answer in step 2; the mean likelihoods were .08, .13, .15, and .13 when the students' answers were error-ridden, vague, partially correct, and completely correct, respectively.

In step 4, the tutor initiates a variety of methods to improve the quality of the answer (see Figure 1). Sometimes the tutor directly splices in the correct answer. More frequently, the tutor uses scaffolding techniques that encourage the student to supply information in a collaborative fashion. For example, the tutor might provide a hint or ask an embedded question, as illustrated below.

> [The tutor and student are discussing how to operationally define the quality of a restaurant.]
>
> TUTOR: What type of scale would that be?
> STUDENT: Oh, let me think, which one. I don't know.
> TUTOR: Try to think. Nominal or (pause)?
> STUDENT: Ordinal, yeah.
> TUTOR: It would be. Why would it be an ordinal scale?

Therefore, the construction of an answer is a collaborative activity -- not a burden that rests entirely on the shoulders of the student. On the average, the tutor ends up supplying more answer information than does the student, even though the tutor originally asks the question (Graesser, 1992).

In step 5, the tutor assesses whether the student understands the answer. In most cases, the tutor simply asks the student whether the student understands ("Do you understand?", "Do you follow?", "Okay?"). Unfortunately, student answers to these comprehension-gauging are inaccurate, as was discussed in the context of claim 3. Tutors occasionally ask a simple follow-up question that tests the student's understanding of the answer (7% of the cases). Very rarely does the tutor thoroughly test the student's understanding by asking a complex question or by requiring the student to solve a problem, as illustrated below.

> TUTOR: Do you have any problem with these kinds of word problems (referring to a section in the book). Where they say--
> STUDENT: (interrupts) Uh, not really.
> TUTOR: You don't? You don't? You don't have any trouble with that?
> STUDENT: No.
> TUTOR: Let's just do one of them. Um, Dan earned 56 dollars, which was twice more than what Jim earns. Now you're supposed to write an equation.
> STUDENT: Uh, I can't write the equations.

Teachers in classrooms normally enact a 3-step dialogue frame instead of a 5-step dialogue frame. Mehan (1979) identified a persistent dialogue pattern in classrooms which includes elicitation, response, and evaluation. The teacher elicits information from the student, the student responds, and then the teacher evaluates the response. This classroom dialogue pattern corresponds to the first three steps of our 5-step dialogue frame in tutoring. What makes tutoring special is the prevalence of the two extra steps (4 and 5). It is conceivable that these extra two steps account for the advantage of tutoring over classroom settings.

Claim 8: Question answering is a collaborative exchange.

Research in conversation has emphasized the point that conversation is a collaborative activity (Clark & Schaefer, 1989; Kreuz & Roberts, 1993). The listener assists the speaker by filling in words and by providing backchannel feedback that acknowledges that the listener is following what the speaker is saying ("uh huh"). The listener does this while the speaker is speaking.

Not surprisingly, question answering is a collaborative activity in tutorial dialogue. This claim is supported in a simple analysis of the number of turns in the answers of tutor questions. There would be only two turns if the student answered the question (step 2) and the tutor supplied feedback (step 3). Mehan's (1979) elicitation-response-evaluation sequence requires a minimum of two turns. In fact, however, there are many more turns when tutors pose questions in a naturalistic tutoring environment. The median number of turns was 5 in the research methods corpus and 10 in the algebra corpus. The tutor and student collaborate in the construction of answers to questions.

Claim 9: Tutors need to pose questions with higher specification.

Tutors elliptically deleted words, phrases, and clauses from their questions under the assumption that the context is sufficiently rich for the student to reconstruct the intended question. Unfortunately, tutors are frequently incorrect in making this assumption. As a consequence, the student ends up misinterpreting the question or answering the wrong question. Tutor questions were classified on degree of specification, with values of high, medium, and low (Graesser & Person, in press). Only 2% of the questions had high specification and 50% had low specification. Students sometimes did not have enough context to interpret the question so they asked counter-clarification questions (see step 1 in Figure 1). The likelihood of a student asking a counter-clarification question decreased as a function of higher question specification, .17, .08, and .00 for tutor questions that were low, medium, versus high in specification. Therefore, tutors should make every effort to formulate their questions with a higher degree of specification.

Claim 10: Tutors need to ask more long-answer questions.

Tutors need to ask better questions in step 1 of the 5-step dialogue frame. More specifically, questions could be posed in a manner that exposes more reasoning on the part of the student, such as the deep reasoning questions. Graesser and Person (in press) reported that there was a tendency for tutors to ask simple short-answer questions that required minimal contributions from the student (e.g., a single word, a YES/NO decision). Tutors need to be trained on question asking skills that encourage the student to become a more substantial contributor.

Claim 11: Tutors need to wait longer for student answers.

Tutors could be more patient in allowing the student to supply an answer in step 2 of the 5-step dialogue frame. Students need time to think, reason, and plan an answer (Dillon, 1988). The knowledge is normally fragile so it takes considerable time to construct an answer. Tutors do frequently pump the student for additional answer information in step 2, as mentioned earlier. However, the tutors could increase the pause duration in step 2 so the student has ample time to think and reason. In a classroom study reported by Swift, Gooding, and Swift (1988), learning improved when teachers increased the pause duration.

Claim 12: The tutor's feedback on student answers needs to be more discriminating.

A good tutor presumably adjusts the feedback in step 3 to the quality of the student's answer in step 2. We performed some analyses that tested this intuitively plausible claim. We segregated student answer contributions into four quality levels: error-ridden, vague (or no answer), partially correct, and completely correct. Short feedback consisted of the brief positive, negative, or neutral responses in step 3 (e.g., "yeah", "right", "good", "okay", "uh huh", "not so", head movement). Long feedback consisted of lengthier comments on answer quality during step 4 (e.g., "that is correct because...", "there is a problem

with your claim that..."). Corrective feedback is a more complex form of negative feedback; the tutor produces information in step 4 that corrects erroneous or misleading information in a student's contribution.

Table 4 presents our analyses of tutor feedback as a function of the quality of the students' contributions. Most of the feedback was provided in the short form (step 3). The long feedback provided a very small increment of evaluative information. Corrective feedback was particularly important in the case of error-ridden answers. We performed statistical analyses on the data by treating each of the 13 tutors from the two corpora as a case. We collapsed the error-ridden and vague answers in order to obtain a sufficient number of observations for each tutor. The likelihood of a tutor giving positive feedback (long or short) increased as a function of answer quality, $F(2,24) = 30.27$, $p < .05$. There were significant differences among all three levels of answer quality (error-ridden/vague, partially correct, versus completely correct). The likelihood of a tutor giving negative feedback significantly decreased as a function of answer quality, $F(2,24) = 24.38$, $p < .05$. Once again, there were significant differences among all pairs of means. These findings indicate that tutors give discriminating feedback to the students.

On the other hand, the tutors were not perfectly discriminating when they administered positive and negative feedback. When error-ridden answers were produced by students, the tutors gave positive and negative feedback with an equal likelihood, $F(1,12) = .01$. When the students produced vague answers, the tutors were more likely to give positive feedback than negative feedback. Clearly, the feedback is off the mark in these cases. Part of the reason for this misleading feedback is that tutors are reluctant to give negative feedback. Perhaps the tutors believe that negative feedback will traumatize the student or reduce the willingness of student to supply information. Alternatively, perhaps tutors are following the politeness conventions of normal conversation (Brown & Levinson, 1987).

Tutors frequently "spliced in" correct information when a student produced error-ridden answers. Yet the tutors did not normally acknowledge the error as an error, or pursue the implications of an error-ridden statement (see also McArthur et al., 1990). There was a significantly higher likelihood of giving corrective feedback than short negative feedback or long negative feedback, $F(2,24) = 35.87$, $p < .05$. It is quite possible that students were unaware that their contributions were error-ridden. Table 5 summarizes how the tutors responded to the errors of the students.

Claim 13: Tutors improve answer quality with a variety of scaffolding strategies.

Step 4 in Figure 1 lists many of the strategies that the tutor uses to improve the quality of the answer. Sometimes the tutor directly splices in the correct answer. Alternatively, the tutor encourages the student to collaborate by asking follow-up questions, giving hints, offering suggestions, and so on. Step 4 is the critical locus of applying scaffolding techniques.

We performed some analyses that traced the evolution of an answer to each question. We observed the quality of contribution N+1, given that the tutor and student had together achieved a particular level of quality via contributions 1 to N. Once again, there were four levels of answer quality: error-ridden, vague/nothing, partially correct, and completely correct. A transition matrix was prepared for the tutor; this specified the likelihood that a tutor supplied a contribution of quality Q at N+1, given that the student and tutor had achieved a cumulative state of quality C at contribution N. A similar transition matrix was prepared for the student. This analysis permitted us to quantify the quality of the information that was supplied by each speech participant.

Table 6 presents the transition matrices for the tutors and students in the two corpora. The data can be interpreted from many perspectives. We were intrigued by three patterns.

A. The tutor waited for the student to supply information when the cumulative quality of the answer was vague or nothing. This generalization can be captured by the following production rule:

IF [quality of cumulative collaborative exchange = vague or no answer]
THEN [tutor pumps student for more information]

The tutors were reluctant to give a completely correct answer when the cumulative quality was vague or no answer; the likelihoods were .12 and .03 in the research methods corpus and the algebra corpus, respectively. The comparable likelihoods for students were significantly higher (.37 and .14). Therefore, it was the student, not the tutor, that supplied correct information in this situation, even though the tutor was more knowledgeable. Tutors normally pumped the student with neutral feedback at step 3 (e.g., "uh huh") in order to encourage the student to supply more information (particularly at the beginning portion of an answer). Tutors were reluctant to rush in with a complete answer at the beginning of the answer evolution.

B. The tutor spliced in a partially correct or completely correct answer when the student committed an error. This generalization is captured by the following production rule:

IF [student's contribution is error-ridden]
THEN [tutor splices in an answer that is partially or completely correct]

The likelihood of a tutor giving a partially or completely correct answer on contribution N+1 significantly varied as a function of the cumulative quality state at contribution N, $F(3,36) = 8.43$, $p < .05$ (when combining the 13 tutors from the two corpora). The likelihoods were .59, .62, .58, and .81 for the quality states of completely correct, partially correct, vague/no-answer, and error-ridden at contribution N. The .81 value was significantly higher than the other values. Therefore, tutors had the tendency to splice in a good answer when students committed errors. They frequently did this without informing the student that the student's answer was error-ridden (see claim 12).

C. The tutor carried the burden of summarizing or recapping the answer. The production rule for this generalization is:

IF [quality of the cumulative collaborative exchange = completely correct]
THEN [tutor supplies a summary or recap of the answer]

Tutors were more likely than students to give a completely correct answer when the cumulative exchange had already reached the quality state of completely correct, .16 versus .04, respectively, $F(1,12) = 6.08$, $p < .05$. It would be preferable for the student to take on the burden of providing these summaries and recaps because such activities improve organization and retention. Tutors perhaps need to be trained to shift this burden onto the student.

There are a large number of sophisticated scaffolding techniques that could be applied in step 4 of the 5-step dialogue frame. Tutors would need to be trained to use these techniques effectively. For example, the modeling-scaffolding-fading technique could be delivered more completely and skillfully. Tutors need to learn how to fade and let the student take more control when they are starting to achieve some success. We were struck by the fragmentary and poorly articulated contributions of the student. As a consequence, the tutors supplied most of the information, leaving the students to fill in short contributions (e.g., a single word, phrase, proposition, step, number). The tutors could relinquish control of the conversation much sooner and could gradually encourage students to supply longer contributions.

Claim 14: Tutors do not adequately assess whether the student understands the answer.

The tutor assesses whether the student understands the answer in step 5 of the 5-step dialogue frame. In 92% of the observations, the tutor conducted this assessment by simply asking the student a comprehension-gauging question (e.g., "Do you understand?", "Do you follow?", "Okay?"). Unfortunately, the students' answers to these comprehension-gauging questions were notoriously unreliable, if not misleading (see claim 3 and Table 1). Tutors apparently assume that students understand anything that gets discussed during tutoring. If something gets said, tutors assume that it must be understood; the tutors merely seek a quick verification from the student that this is the case.

A good tutor would assess the student's understanding more rigorously. The tutor could ask one or more follow-up questions that are diagnostically discriminating and that troubleshoot potential misunderstandings. The tutor could present a similar problem and request that the student solve it in order

to actively demonstrate understanding. However, the 13 tutors in our naturalistic sample were rarely rigorous in step 5.

Claim 15: Tutors need to violate some pragmatic rules of polite conversation.

The pragmatic rules of normal polite conversation have been identified by Grice (1975) and others (Brown & Levinson, 1987). These rules are pervasive and highly automatized. Unfortunately, they sometimes present a barrier to effective pedagogy. A good tutor may need to violate some rules and conversational maxims in order to crack the barrier. For example, rather than following the Gricean "maxim of quantity," tutors need to be redundant and repetitious to enhance student understanding. Instead of being polite and "face saving" when a student makes an error, the tutor needs to "take off the gloves" and directly confront the student.

The rules followed by participants in normal conversations have been described by Grice (1975). Discourse is governed by one overarching cooperative principle: conversational participants make a good faith effort to contribute and to collaborate in the ongoing discourse. Cooperation is augmented by four conversational maxims: quantity (don't say more or less than is required), quality (don't say things that are untrue or that lack evidence), relevance (don't say things that are extraneous), and manner (don't say things that are vague or disordered).

Brown and Levinson (1987) studied linguistic politeness in several cultures. They proposed some general principles and discourse strategies to facilitate social interaction. Central to their analysis is the notion of a face, or one's self image. Individuals in a culture attempt to maintain a positive self-image, and help others to maintain their self-images. This is not always possible, however, because face is frequently endangered by face threatening acts, such as requests, criticisms, and demands. Each culture has a number of linguistic strategies to mitigate the impact of these face-threatening acts.

Table 7 presents some of the maxims of Grice and politeness strategies of Brown and Levinson. Associated with each of these are costs and benefits from the perspective of effective pedagogy during tutoring. It is appropriate to follow the maxims and politeness strategies under some conditions, but to violate them under other conditions.

The following example illustrates that there are potential pedagogical costs to the politeness strategy of "avoiding disagreement." The tutor and student were discussing various types of graphs.

> TUTOR: ...and that's our frequency distribution... What is that one called again (pointing to a bar graph)?
> STUDENT: A histogram.
> TUTOR: Alright, or a bar graph.
> STUDENT: Bar graph.

The student failed to acknowledge the important distinction between histograms (involving continuous variables) and bar graphs (involving discrete variables). However, the tutor did not acknowledge that the student had made an error; in fact, the tutor gave potentially positive feedback in step 3 ("alright"). The tutor was sufficiently ambiguous in step 4 to permit the erroneous interpretation that a histogram and a bar graph are interchangeable.

Once again, a good tutor may need to breach the normal conversational maxims and politeness strategies. This could be very uncomfortable to the student, of course. A possible solution to this problem would be to establish some "conversational ground rules" at the beginning of a tutoring session. The tutor could explain to the student that it is important for the tutor to provide critical feedback, to point out misconceptions, and to challenge the student. The tutor could encourage the student to articulate answers in detail and not to get rattled when negative feedback is given. The tutor could resurrect the adage that students learn from their errors. It is a question for further research whether these conversational ground rules will minimize face-threatening acts during tutoring, and whether systematic violations of maxims will facilitate learning.

Computational Models of Speech Act Prediction: Quantification of Dialogue Patterns

Researchers in discourse processing, sociology, and sociolinguistics have analyzed prominent dialogue patterns (Clark & Schaefer, 1989; D'Andrade & Wish, 1985; Goffman, 1974; Graesser, 1992; Mehan, 1979; Schegloff & Sacks, 1973; Turner & Cullingford, 1989). Some of the systematicity resides at a categorical level that does not consider the world knowledge, beliefs, and goals of the speech participants. That is, there are appropriate orderings of speech act categories and inappropriate orderings. Schegloff and Sacks (1973) analyzed the adjacency pairs of conversational turns: Given that one speaker utters a speech act in category C during turn N, what is the appropriate speech act category for the other speaker at the next, adjacent turn N+1? The most common adjacency pair is the [Question --> Reply-to-question] sequence. The adjacency pair analysis considers only one speech act of prior context when generating predictions for the subsequent speech act.

Researchers have identified larger sequences of dialogue patterns. Mehan (1979) identified a frequent triple in classroom environments, as illustrated below.

> TEACHER QUESTION: What is the capital of Florida?
> STUDENT ANSWER: Athens.
> TEACHER EVALUATION OF ANSWER: No, that's not right.

As discussed in the previous section, this triplet is expanded to a 5-step dialogue frame in tutoring environments. Counter-clarification questions produce a quadruple sequence, as illustrated below.

> QUESTION-A: Where did you go yesterday?
> QUESTION-B: Yesterday morning?
> ANSWER-B: Yeah, in the morning.
> ANSWER-A: To Jack's, for breakfast.

The knowledge accumulated in the study of dialogue patterns has been fragmented and largely untested. No one has developed a model that ties together the assorted observations. No one has quantified how successfully these patterns account for the speech acts in naturalistic conversation. There is no model that is sufficiently broad in scope that it could be applied to any conversation or text. In view of these shortcomings, we developed some computational models that attempt to capture the systematicity in speech act sequences (Graesser, Swamer, Baggett, & Sell, in press; Swamer, Graesser, Franklin, Sell, Cohen, & Baggett, 1993). Two classes of the models have radically different computational architectures: a connectionist architecture and a symbolic architecture.

The computational models assume that the stream of conversation (or text) can be segmented into a linear sequence of speech act categories. There have been extensive debates over what speech act categories are needed for a satisfactory analysis of human conversation (see D'Andrade & Wish, 1985). We adopted a slightly modified version of D'Andrade and Wish's (1985) set of speech act categories. Their categories were both theoretically motivated and empirically adequate in the sense that trained judges could agree on the assignment of categories. Table 8 presents the 8 speech act categories that were adopted in our analyses. Given that there are two speakers in a dialogue, each speech act in a conversation can be in one of 16 categories (2 speakers x 8 basic speech acts = 16). A Juncture (J) category was also included in order to signify lengthy pauses in a conversation and excerpts that are uninterpretable to judges. This yielded 17 categories altogether. In summary, the stream of dyadic conversation was segmented into a sequence of speech acts and each speech act was assigned to one of 17 speech act categories.

Conversations analyzed

Children's' dyads. Sell, Cohen, Crain, Duncan, MacDonald, and Ray (1991) adopted this 17-category speech act scheme in their analysis of 90 conversations involving pairs of children. Dyads of second graders and sixth graders were videotaped for 10 minutes in three different contexts: playing 20 questions, solving of a puzzle, and free play. The dyads were further segregated according to how well they knew each other: mutual friends (A and B like each other), unilateral friends (A likes B, but B neither likes nor

dislikes A), and acquaintances (A and B do not like each other or dislike each other). All of the children in the dyads were from the same classroom so they were never strangers. Sell et al. (1991) reported that the 17-category speech act scheme could be successfully applied to the 16,657 speech acts in this corpus. Trained judges could segment the stream of conversation into speech acts with high reliability. The 17 categories were sufficiently complete in the sense that all of the speech acts fit into one of the 17 categories. Trained judges also could reliably categorize the speech acts; the Cohen's kappas were .82, .76, and .74 for the question task, the puzzle task, and the free play task, respectively. There was a mean of 2.3 speech acts per conversational turn.

College tutoring. A subset of the research methods tutoring corpus was extracted and analyzed. We extracted all deep reasoning questions posed by the tutor (i.e., why, how, what-if, as discussed earlier). The question and answer sequence for each of these questions was included in the college tutoring corpus. There were 2013 speech acts in this corpus, and a mean of 2.9 speech acts per conversational turn.

Telephone conversations. We had access to a corpus of telephone conversations recorded by the Nynex corporation. The conversations were between telephone operators and customers in New York City. There were 1102 speech acts in this corpus, and 2.5 speech acts per turn.

Goodness-of-prediction (GOP) score

The goal of each model was to capture the systematicity in the sequential ordering of the speech act categories. That is, to what extent can the category of speech act N+1 be successfully predicted, given the sequence of speech acts 1 through N? A hit rate is the likelihood that a theoretically predicted category actually occurs in the data, as specified in formula 1.

$$p(\text{hit}) = p(\text{category C occurred at N+1} \mid \text{category C is predicted by the model at N+1}) \quad (1)$$

A hit rate is not a satisfactory index of the success of a model, however, because there is no consideration of the likelihood that a speech act would occur by chance. For example, if a particular speech act category occurred in the corpus 90% of the time, then there would be a high hit rate, assuming that the model predicted that category most of the time. A satisfactory index of the model's success would need to control for the baserate likelihood that the predicted speech act occurred in the empirical distribution of speech act categories (called the a posteriori distribution). For example, the baserate likelihoods of the speech act categories in the Sell corpus were .21, .14, .04, .02, .40, .03, .07, .03, and .07 for categories Q, RQ, D, ID, A, E, R, N, and J, respectively. We computed a goodness-of-prediction (GOP) score that corrected for the baserate likelihood that a speech act category would occur by chance, as specified in formula 2.

$$\text{GOP score} = [\text{hit-rate(category C)} - \text{baserate(C)}]/[1.0 - \text{baserate(C)}] \quad (2)$$

Sometimes a model specified that more than one speech act category could occur at observation N+1. In this case, formulas 1 and 2 are still correct except that the values are based on a set of categories rather than a single category.

Computational models

Recurrent connectionist network. Researchers in the connectionist camp of cognitive architectures have developed a recurrent network that is suitable for capturing the systematicity in the temporal ordering of events (Cleeremans & McClelland, 1991; Elman, 1990). The recurrent connectionist network preserves an encoding of all previous input, and uses this information to induce the structure underlying temporal sequences.

There are four layers of nodes in the recurrent network, as shown in Figure 2. The input layer specifies the category of speech act N. There are 17 nodes in the input layer, one for each speech act category. The appropriate node is activated when speech act N is received. For example, if person 1 asked a question, then the Q1 node would be activated in the input layer of the network. The output layer contains the network's predictions for speech act N+1. There are 17 output nodes, one for each speech act category.

An output node has an activation value that reflects the degree to which the network predicts that output node. If the input were Q1, for example, then we would expect RQ2 to receive a high activation value in the output layer. This would capture the regularity that people are expected to answer questions that others ask. The hidden layer captures higher order constituents that are activated by speech act N. Hidden layers are frequently implemented in connectionist architectures in order to capture internal cognitive mechanisms (Rumelhart & McClelland, 1986). The hidden layer is needed when direct input-output mappings fail to capture systematicity in the data. There were 10 nodes in the hidden layer of our network. The context layer allows the network to induce temporal sequences. The context layer stores the activations from the hidden layer of the previous step in the speech act sequence (as designated by the fixed weights of 1 in Figure 2). The activations of the hidden layer at step N depend on: (a) the input at N and (b) the activation of the context layer at N (which was the hidden layer at N-1). Therefore, the hidden layer is receiving information about the present input and past inputs. The resulting activation pattern of the hidden layer's 10 nodes at step N is subsequently copied into the context layer at step N+1. The context layer must have the same number of nodes as the hidden layer, namely 10 nodes in our model.

There are a total of 440 connections that are allowed to vary in the weight space of this model. There are 170 connections between the input layer and the hidden layer, given that there are 17 input nodes and 10 hidden layer nodes. Similarly, there are 170 connections from the hidden layer to the output layer. The other 100 nodes link the 10-node context layer to the 10-node hidden layer. There are also connections from the hidden layer to the context layer that are fixed at 1.0. In preliminary simulations, we varied the number of nodes in the hidden layer and the context layer (from 6 to 14 nodes). However, the success of the model did not significantly depend on the number of nodes in these layers, at least within the range of 6 to 14 nodes.

The performance of the recurrent network was evaluated by computing two different GOP scores (see formula 2). A maximal activation GOP score considered only one output node as the predicted speech act category for step N+1. The predicted category was the one that had received the highest activation value in the output layer. An above-threshold GOP score allowed for the network to accommodate multiple speech act categories at each step. All output nodes that met or exceeded a threshold activation level were predictions for step N+1. Preliminary tests had revealed that a threshold of .18 provided an appropriate fit to the three corpora. On the average, 1.7 speech acts were above threshold at any given step in the conversation.

We tested some connectionist models that removed one or more components of the recurrent connectionist model. This permitted us to assess which components of the recurrent connectionist model had the most robust impact on the prediction of speech act systematicity.

Double-entry backpropagation network. This network considered only two speech acts of context (N-1 and N) when predicting speech act N+1. This was accomplished by removing the context layer of the recurrent network (see Figure 2) and adding 17 nodes for N-1 as additional nodes in the input layer (yielding 34 input nodes). The hidden layer was preserved. There were 510 connections in the weight space for this network.

Single-entry backpropagation network. This network considered only one speech act of context (N) when predicting speech act N+1. This was accomplished by removing the context layer of the recurrent network, but preserving the hidden layer. There were 340 connections in the weight space.

Perceptron. This network removed both the hidden layer and the context layer of the recurrent network. Thus, there were direct connections between the input layer and the output layer. There were 289 connections in the weight space (17 x 17 = 289).

Recursive transition network (RTN) . This model had a symbolic computational architecture (Graesser, Swamer, Baggett, & Sell, in press; Stevens & Rumelhart, 1975). One advantage of a symbolic architecture is that the investigator can trace and articulate the dialogue patterns that explain systematicity in the data. In contrast, it is difficult to identify patterns in a weight space from a connectionist model and to articulate the patterns succinctly.

Figure 3 shows a recursive transition network (RTN) for speech act prediction that was developed by Graesser, Swamer, Baggett, and Sell (in press). Some modules in the RTN would be anticipated on the basis of common sense and theoretical developments in the literature. Following Clark and Schaefer (1989), for example, the RTN in Figure 3 segregates a Contribution from an Acknowledgment of the contribution by the other party. There are four modules that emanate from the Contribute node (Interrogate, Inform, Direct, and Evaluate), which capture four basic goals of communication. Counter-clarification questions (i.e., k-Interrogate) are embedded in the second step of the Interrogate, Direct, and Evaluation modules. The Challenge module is a reaction of person A when person B tries to evaluate something or B tries to get A to do something (i.e., the Direct and Evaluate modules, respectively).

The RTN in Figure 3 has seven modules, altogether. Each module has two or three state nodes and a set of arcs that emanate from each state node. The arc specifies the set of legal speech act categories and set of recursively embedded modules that are legal at that point. The speech act categories are the same 8 categories that were defined earlier: Q, RQ, D, ID, A, E, V, and N. There are 7 recursively embedded modules: Contribute, Acknowledge, Interrogate, Direct, Evaluate, Inform, and Challenge. The i and k are indices that keep track of which of the two individuals is speaking. In some cases, the same individual produces a sequence of speech acts. In other cases, the turn transfers to the other person.

The RTN generates a set of legal speech acts at each step of the conversation. A speech act at N+1 is legal if there is at least one path in the family of alternative paths that emanate from speech act N. A hit occurs when speech act N+1 matches one of the legal alternatives. Hit rates and GOP scores can be computed in the same way that they were computed for the recurrent connectionist network (see formulas 1 and 2). In a discrete RTN, there is an all-or-none prediction for each speech act at step N+1. In a weighted RTN, each arc is weighted according to the likelihood that the arc would be traversed while accounting for the speech act corpus; consequently, each speech act was predicted with some likelihood that varied from 0 to 1. We tested a weighted RTN because it provided a closer fit to the data. This was accomplished by an optimization procedure that determined the best-fit set of weights which maximized the GOP score. A speech act was scored as predicted if it met or exceeded a strength threshold.

Schegloff and Sacks' adjacency network. This was an RTN that captured the adjacency pair analysis of Schegloff and Sacks (1973). Therefore, only one speech act of context would be considered when predicting speech act N+1, and the speaker of N was always a different speaker than the speaker of N+1. The speech act categories of Schegloff and Sacks were translated into those categories in Table 8.

Performance of models in predicting speech act categories

Table 9 presents performance data on the four connectionist models of speech act prediction. Goodness-of-prediction (GOP) scores are listed for each model and corpus. Table 9 also includes the hit rate, baserate, and mean number of speech acts predicted by the recurrent connectionist network. It was possible to perform statistical analyses on the simulations of the connectionist networks by having a different set of random starting weights in the weight space and running the simulation 10 times. As a crude, but conservative estimate, a GOP score difference of .010 is significant ($p < .05$).

Maximum activation GOP scores were available for the four connectionist models. The predicted speech act for a model was the one speech act that had the highest activation score. The recurrent connectionist network was the best network according to this performance measure. When averaging over the three corpora, the GOP scores were .337, .317, .290 and .290 for the recurrent network, the double-entry backpropagation network, the single-entry backpropagation network, and the perceptron. A very similar pattern of scores emerged for the above threshold GOP scores, where more than one speech act was predicted: .439, .442, .326, and .328, respectively. In this case, however, there was no difference between the recurrent network and the double-entry backpropagation network. These results are consistent with the following conclusions.

1. The recurrent connectionist network correctly predicts the next speech act 34-44% of the time (after controlling for baserate guessing).

2. The average number of predicted speech act categories is 1.7.

3. Only 2 (or possibly 3) speech acts of context are effective in formulating successful predictions of the next speech act category. (This was further substantiated in follow-up analyses of the recurrent network that plotted GOP scores as a function of the number of context items available).

4. Two speech acts of context are much better than one.

The third conclusion suggests that it is futile for speakers to plan several speech acts into the future. Speakers are constantly replanning, re-evaluating, and revising the conversation in the face of constantly changing situational constraints (Clark & Schaefer, 1989; McArthur et al., 1990; Winograd & Flores, 1986). Speaker A's next speech act category appears to be formulated on the basis of speaker A's last speech act, together with speaker B's last speech act. The context prior to this is not very useful for formulating predictions. A global, top-down, expectation-driven model of conversation would have problems explaining our results.

The performance on the recurrent connectionist network was compared to the two recursive transition networks. In order to compare each RTN network with the recurrent connectionist network, we computed a model comparison ratio, which is specified in formula 3.

$$\text{Ratio} = \text{GOP (RTN} \mid \text{S speech acts predicted)} / \text{GOP (recurrent} \mid \text{S speech acts predicted)} \qquad (3)$$

The GOP score of the recurrent network was yoked to the GOP score of the RTN network so that both models predicted the same number of speech acts at N+1 (on the average). A model comparison ratio score of 1 means that the two models perform the same. A ratio of less than 1 means the recurrent network performs best, whereas a ratio of greater than 1 means that the RTN performs best.

The recurrent connectionist network performed better than the two RTN's. The maximum values of the model comparison ratios were determined over varying values of S (i.e., number of predicted speech acts, which vary with the threshold value). For Graesser's RTN, the maximum values were .89, .43, and .50 in the children's dyad corpus, the college tutoring corpus, and the telephone corpus, respectively. The mean number of predicted speech acts at a step were 6.6, 2.9, and 3.7, respectively. Therefore, on the average, 61% of the systematicity that was picked up by the recurrent connectionist network was also captured by Graesser's RTN. The performance of the Schegloff and Sacks RTN was much worse. The maximum model comparison ratios were .53, .29, and .12, respectively, so this second RTN captured only 31% of the systematicity of the recurrent connectionist network. In this case, the mean numbers of predicted speech acts at a step were 2.7, 2.8, and 2.9, respectively. The fact that the adjacency RTN performed much more poorly than the Graesser RTN supports conclusion 4 (i.e., two speech acts of context are quite a bit better than one).

Viewed from another perspective, it could be argued that Graesser's RTN did an impressive job in capturing the systematicity of the speech act sequencing. We might view the recurrent connectionist model as a statistical upperbound in capturing the sequential systematicity in dialogue patterns (when considering only speech act categories, not the content of the speech acts). Graesser's RTN captures 61% of the upperbound in systematicity. This is perhaps an impressive figure.

Additional analyses

Follow-up analyses were performed in order to answer some additional questions about the dialogue patterns. We analyzed the children's dyad data to assess whether GOP scores varied as a function of type of task, age, and type of relationship. These analyses revealed that the type of task had a robust impact on GOP scores. The maximum activation GOP scores were .38, .07, and .18 in the question task, the puzzle

task, and free play, respectively. The children apparently engaged in parallel monologues in the puzzle task, whereas the 20-questions game placed substantial constraints on the dialogues. In contrast, the age of the children and the type of social relationship (i.e., mutual friends, unilateral friends, versus acquaintances) had absolutely no impact on the GOP scores.

References

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. Cognitive Science, 13, 467-505.

Anderson, J. R., & Reiser, B. J. (1985). The LISP Tutor. Byte, 10, 159-175.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives. Handbook I: Cognitive domain. New York: McKay.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 4-16.

Bransford, J. D., Arbitman-Smith, R., Stein, B. S., & Vye, N. J. (1985). Analysis--Improving thinking and learning skills: An analysis of three approaches. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), Thinking and learning skills (Vol. 1, pp. 133-206). Hillsdale, NJ: Erlbaum.

Bransford, J. D., Goldman, S. R., & Vye, N. J. (1991). Making a difference in people's ability to think: Reflections on a decade of work and some hopes for the future. In R. J. Sternberg & L. Okagaki (Eds.), Influences on children. Hillsdale, NJ: Erlbaum.

Brown, A. L. (1988). Motivation to learn and understand: On taking charge of one's own learning. Cognition and Instruction, 5, 311-321.

Brown, J. S., & Levinson, S. C. (1987). Politeness: Some universals in language use. Cambridge: Cambridge University Press.

Burton, R. R., & Brown, J. S. (1982). An investigation of computer coaching for informal learning activities. In D. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems (pp. 79-98). New York: Academic Press.

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145-182.

Clancey, W. J. (1983). GUIDON. Journal of Computer-Based Instruction, 10, 8-15.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. Cognitive Science, 13, 259-294.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. Journal of Experimental Psychology: General, 120, 235-253.

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. American Educational Research Journal, 29, 573-604.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. American Educational Research Journal, 19, 237-248.

Collins, A. (1985). Teaching reasoning skills. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), Thinking and learning skills (Vol. 2, pp. 579-586.) Hillsdale, NJ: Erlbaum.

Collins , A. (1988). Different goals of inquiry teaching. Questioning Exchange, 2, 39-45.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 453-494). Hillsdale, NJ: Erlbaum.

D'Andrade, R. G., & Wish, M. (1985). Speech act theory in quantitative research on interpersonal behavior. Discourse Processes, 8, 229-259.

Dillon, J. T. (1988). Questioning and teaching: A manual of practice. New York: Teachers College Press.

Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A Component analysis. Journal of Educational Psychology, 84, 331-339.

Fishbein, H. D., Eckart, T., Lauver, E., Van Leeuwen, R., & Langmeyer, D. (1990). Learners' questions and comprehension in a tutoring setting. Journal of Educational Psychology, 82, 163-170.

Fitz-Gibbon, C. T. (1977). An analysis of the literature of cross-age tutoring. Washington, DC: National Institute of Education, (ERIC Document Reproduction Service No. ED 148 807).

Fox, B. (1992). Human tutorial dialogue. (Final Report). University of Colorado, Institute of Cognitive Science.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. Memory & Cognition, 10, 597-602.

Goffman E. (1974). Frame analysis. Cambridge, MA: Harvard University Press.

Graesser, A. C. (1992). Questioning mechanisms during complex learning. Memphis State University, Memphis, TN, (ERIC Document Reproduction Service No. ED 350 306).

Graesser, A. C. (1993). Dialogue patterns during naturalistic tutoring. In Proceedings of the 15th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Graesser, A. C., & Franklin, S. P. (1990). QUEST: A cognitive model of question answering. Discourse Processes, 13, 279-303.

Graesser, A. C., Gordon, S.E., & Brainerd, L.E. (1992). QUEST: A model of question answering. Computers and Mathematics with Applications, 23, 733-745.

Graesser, A. C., & Hemphill, D. (1991). Question answering in the context of scientific mechanisms. Journal of Memory and Language, 30, 186-209.

Graesser, A. C., Lang, K. L., & Roberts, R. M. (1991). Question answering in the context of stories. Journal of Experimental Psychology: General, 120, 254-277.

Graesser, A. C., Langston, M. C., & Baggett, W. B. (1993). Exploring information about concepts by asking questions. In G. V. Nakamura, R. M. Taraban, and D. Medin (Eds.), Acquisition, representation, and processing of categories and concepts. Orlando, FL: Academic Press.

Graesser, A. C., Langston, M. C., & Lang, K. L. (1992). Designing educational software around questioning. Journal of Artificial Intelligence in Education, 3, 235-243.

Graesser, A. C., & McMahen, C. L. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. Journal of Educational Psychology, 85, 136-151.

Graesser, A. C., & Person, N. K. (in press). Question asking during tutoring. American Educational Research Journal.

Graesser, A. C., Person, N. K., & Huber, J. (1992). Mechanisms that generate questions. In T. Lauer, E. Peacock, & A. Graesser (Eds.), Questions and information systems. Hillsdale, NJ: Erlbaum.

Graesser, A. C., Person, N. K., & Huber, J. (1993). Question asking during tutoring and in the design of educational software. In M. Rabinowitz (Ed.), Cognitive science foundations of instruction (pp. 149-172). Hillsdale, NJ: Erlbaum.

Graesser, A. C., Swamer, S. S., Baggett, W. B., & Sell, M. A. (in press). New models of deep comprehension. In B. K Britton & A. C. Graesser (Eds.), Models of understanding text. Hillsdale, NJ: Erlbaum.

Greeno, J. (1982). Forms of understanding in mathematical problem solving. In S. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), Learning and motivation in the classroom. Hillsdale, NJ: Erlbaum.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and Semantics, Vol 3: Speech acts (pp. 41-58). New York: Academic Press.

Kerry, T. (1987). Classroom questions in England. Questioning Exchange, 1 (1), 32-33.

Kieras, D.E. (1992). Explanation-based knowledge acquisition of electronics: Final report. Technical Report No. 36, Technical Information Design and Analysis Laboratory, University of Michigan, Ann Arbor, Michigan.

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. Contemporary Educational Psychology, 14, 366-381.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. American Educational Research Journal, 29(2), 303-323.

Kreuz, R. J., & Roberts, R. M. (1993). When collaboration fails: Consequences of pragmatic errors in conversation. Journal of Pragmatics, 19, 239-252.

Leinhardt, G. (1987). Development of an expert explanation: An analysis of a sequence of subtraction lessons. Cognition and Instruction, 4, 225-282.

Lesgold, A. (1992). Going from intelligent tutors to tools for learning. In C. Frasson, G. Gauthier, & G. I. McCalla (Eds.), Intelligent tutoring systems. Berlin: Springer.

Mayer, R. (1992). Cognition and instruction: Their historic meeting within educational psychology. Journal of Educational Psychology, 84, 405-412.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. Cognition and Instruction, 7, 197-244.

Miyake, N., & Norman, D. A. (1979). To ask a question, one must know enough to know what is not known. Journal of Verbal Learning and Verbal Behavior, 18, 357-364.

Mehan, H. (1979). Learning lessons: Social organization in the classroom. Cambridge, MA: Harvard University Press.

Mohan, M. (1972). Peer tutoring as a technique for teaching the unmotivated. Fredonia, NY: State University of New York, Teacher Education Research Center. (ERIC Document Reproduction Service No. ED 061 154).

Moore, J.D. & Ohlsson, S. (1992). Educating migraine patients through on-line generation of medical explanations. Proceedings of the 14th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. Cognition and Instruction, 9, 329-390.

Ohlsson, S. (1986). Some principles of intelligent tutoring. Instructional Science, 14, 293-326.

Ohlsson, S., & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. Cognition and Instruction, 8, 103-179.

Palincsar, A. S., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 1 (2), 117-175.

Papert, S. (1980). Mindstorms: Children, computers and powerful ideas. New York: Basic Books.

Person, N. K, Graesser, A. C., Magliano, J. P., & Kreuz, R. J. (1993). Inferring what the student knows in one-to-one tutoring: The role of student questions and answers. Manuscript under editorial review. Memphis State University, Memphis, TN.

Pressley, M. (1990). Cognitive strategy instruction that really improves children's academic performance. Cambridge, MA: Brookline Books.

Pressley, M., Symons, S., McDaniel, M. A., Snyder, B. L., & Turnure, J. E. (1988). Elaborative interrogation facilities in the acquisition of confusing facts. Journal of Educational Psychology, 80, 301-342.

Putnam, R. T. (1987). Structuring and adjusting content for students: A live and simulated tutoring of addition. American Educational Research Journal, 24, 13-48.

Reiser, B.J., Kimberg, D.Y., Lovett, M.C., & Ranney, M. (1991). Knowledge representation and explanation in GIL, an intelligent tutor for programming. In J.H. Larkin and R.W. Chabay (Eds.), Computer assisted instruction and intelligent tutoring systems: Shared issues and complementary approaches (pp. 111-149). Hillsdale, NJ: Erlbaum.

Rogoff, B. (1990). Apprenticeship in thinking. New York: Oxford University Press.

Rosenshine, B., & Chapman, S. H. (1990). Teaching students to generate questions: A review of different facilitators. Unpublished manuscript, University of Illinois, Champaign-Urbana.

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 1). Cambridge: MIT Press.

Sanders, G. A., Evens, M. W., Hume, G. D., Rovick, A. A., & Michael, J. A. (1991). An analysis of how students take the initiative in keyboard-to-keyboard tutoring dialogues in a fixed domain. Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society (pp. 1086-1091). Hillsdale, NJ: Erlbaum.

Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. Cognition and Instruction, 9, 177-199.

Scardamalia, M., Bereiter, C., McLean, R. S., Swallow, J., & Woodruff, E. (1989). Computer supported intentional learning environments. Journal of Educational Computing Research, 5, 51-68.

Schank, R., Ferguson, W., Birnbaum, L., & Greising, M. (1991). ASK TOM: An experimental interface for video case libraries. The Proceedings of the 13th Annual Conference for the Cognitive Science Society (pp. 570-575). Hillsdale, NJ: Erlbaum.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. Semiotica, 8, 289-327.

Sebrechts, M. M., & Swartz, M. L. (1991). Question-asking as a tool for novice computer skill acquisition. Proceedings of the International Conference on Computer-Human Interaction, 293-297.

Sell, M. A., Cohen, R., Crain, M., Duncan, M. K., MacDonald, C. D., & Ray, G. E. (1991, April). The context of dyadic interactions: Communications as a function of task demands and social relationships. Society for Research on Child Development, Seattle, WA.

Singer, M., & Donlan, D. (1982). Active comprehension: Problem solving schema with question generation for comprehension of complex short stories. Reading Research Quarterly, 17, 166-186.

Sleeman, D. H., & Brown, J. S. (Eds.) (1982). Intelligent tutoring systems. New York: Academic Press.

Stevens, A. L., & Rumelhart, D. E. (1975). Errors in reading: Analysis using an augmented transition network model of grammar. In D. Norman and D. Rumelhart (Eds.), Explorations in cognition. San Francisco: Freeman.

Swamer, S. S., Graesser, A. C., Franklin, S. P., Sell, M. A., Cohen, R., & Baggett, W. B. (1993). A connectionist model of speech act prediction. Proceedings of the 15th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Swift, J. N., Gooding, C. T., & Swift, P. R. (1988). Questions and wait times. In J. T. Dillon (Ed.), Questioning and discussion: A multidisciplinary study. Norwood, NJ: Ablex Publishing Corporation.

Turner, E. H., & Cullingford, R. E. (1989). Using conversation MOPs in natural language interfaces. Discourse Processes, 12, 63-90.

Van Lehn, K. (1990). Mind bugs: The origins of procedural misconceptions. Cambridge, MA: MIT Press.

Weaver, C. A. III (1990). Calibration and assessment of comprehension. Unpublished Doctoral Dissertation, University of Colorado, Boulder, CO.

Winograd, T., & Flores, C. F. (1986). Understanding computers and cognition. Norwood, NJ: Ablex.

Wong, B. Y. L. (1985). Self-questioning instructional research: A review. Review of Educational Research, 55, 227-268.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. American Educational Research Journal, 29, 663-676.

Table 1

Correlations Between Student Achievement and Properties of Student Questions and Answers

Measures of Student Questions and Answers	Achievement Measure	
	Examination Scores	Final Grade
Total number of student questions	-.22	-.34***
Proportion of student questions that are knowledge deficit questions	.15	.32
Proportion of student questions that are deep-reasoning questions	.44*	.58*
Proportion of students' answer contributions that are:		
Completely correct	.32**	.43*
Partially correct	.09	-.09
Vague or no answer	-.30	-.46*
Error-ridden	-.32**	-.10
Error-ridden, vague, or no answer	-.52*	-.49*
Proportion of **Yes** answers (by student) to comprehension-gauging questions (by tutor)	.07	.05
Proportion of **No** answers (by student) to comprehension-gauging questions (by tutor)	.42*	.20

* $p < .05$, two-tailed

** $p < .06$, one-tailed

*** $p < .10$, two-tailed

Table 2

Mechanisms that Generate Tutor Questions.

	CORPUS	
MECHANISMS	Research Methods	Algebra
Curriculum script	.70	.93
Driven by student error	.05	.06
Elaboration of an idea	.19	.03
Summary-recap	.14	.01
Get student to justify something, explain something, or generate an example	.14	.01
Other	.03	.00

Table 3

Continuations After Tutor Question Is Answered.

	CORPUS	
	Research Methods	Algebra
Activity or question guided by tutor's curriculum script	.67	.79
Tutor diagnosis, dissects, or remediates student errors	.02	.04
Elaboration of an idea	.22	.03
Summary - recap	.15	.06
Tutor prompts student to introduce next topic or example	.05	.00
Student initiates next topic or example	.05	.10
Other	.05	.01

Table 4
Tutor Feedback as a Function of Quality of Student Contributions.

		Quality of Student Answer			
Measure	Corpus	Error-ridden	None or Vague	Partially Correct	Completely Correct
Number of observations	Research Methods	48	56	131	130
	Algebra	47	13	109	25
Proportion of observations	Research Methods	.13	.15	.36	.36
	Algebra	.24	.07	.56	.13
Positive Feedback					
Short feedback	Research Methods	.31	.40	.47	.56
Long or short feedback	Research Methods	.31	.45	.50	.63
Short feedback	Algebra	.30	.23	.65	.80
Long or short feedback	Algebra	.30	.31	.73	.92
Negative feedback					
Short feedback	Research Methods	.10	.00	.01	.00
Long or short	Research Methods	.12	.04	.03	.00
Long, short, or corrective	Research Methods	.40	.12	.07	.04
Short feedback	Algebra	.36	.15	.10	.00
Long or short	Algebra	.36	.15	.11	.00
Long, short, or corrective	Algebra	.83	.23	.17	.04
Neutral or No feedback					
	College	.31	.50	.44	.33
	Algebra	.11	.54	.12	.08

Table 5

Analysis of Student Errors Manifested in the Sample of Tutor Questions.

	CORPUS	
	Research Methods	Algebra
Number of errors in sample	48	47
Type of error		
Slip	.16	.13
Bug or glitch	.25	.23
Deep misconception	.59	.63
Tutor's treatment of error		
Error is acknowledged in short or long feedback	.12	.36
Tutor splices in correct answer	.40	.36
Tutor supplies a hint	.10	.45
Tutor reasons to expose derivation of correct answer	.17	.34
Tutor asks student question to extract correct answer	.17	.21
Tutor issues directive to extract correct answer	.04	.06
Likelihood of the student catching his/her own error	.00	.04

Table 6

Contribution Transition Matrix: Status of Contribution on Turn N+1, Given the Cumulative Quality of the Answer During Turns 1 to N.

TUTOR CONTRIBUTION

Research methods corpus

Turn N	Turn N+1 E	N/V	PC	CC
CC	.00	.59	.24	.17
PC	.00	.46	.39	.14
N/V	.00	.56	.32	.12
E	.06	.21	.44	.27

Algebra corpus

Turn N+1 E	N/V	PC	CC
.00	.26	.60	.14
.00	.29	.62	.08
.00	.28	.69	.03
.00	.10	.78	.12

STUDENT CONTRIBUTION

Research methods corpus

Turn N	Turn N+1 E	N/V	PC	CC
CC	.01	.76	.19	.04
PC	.08	.54	.25	.14
N/V	.09	.33	.21	.37

Algebra corpus

Turn N+1 E	N/V	PC	CC
.04	.68	.24	.03
.12	.48	.34	.06
.21	.27	.38	.14

CC = Completely correct answer
PC = Partially correct answer
N/V = Nothing or vague answer
E = Error-ridden answer

Table 7

How conversational rules and politeness strategies may affect feedback during tutoring.

Type	Benefits	Costs
Gricean maxims		
Maxim of Quantity (don't make contribution more informative than is required)	Student is not bored by repetition of material that has been mastered	Tutor may need to be redundant and repetitious to ensure understanding by the student
Maxim of Quality (don't say what you believe to be false or lack evidence)	Student is not confused by contradictory or incorrect statements	Tutor may need to expand on a student's error in order to expose the student's misconceptions (as in Socratic tutoring)
Maxims of Manner (be brief and orderly)	Student benefits from a systematic exposition of the material	Tutor may need to backtrack or skip over topics to identify or correct a student's weaknesses
Brown and Levinsnon's Politeness Strategies		
Exaggerate (show interest, and approval of hearer)	Student is reinforced for mastering the material	Feedback may be inappropriate for the response (e.g., ignoring minor errors in the response)
Understate (allows multiple interpretations)	Softens the blow of critical feedback	Feedback may be inappropriate for the response (e.g., minimizing errors even when they are severe)
Avoid disagreement	Student is not challenged or put in adversarial relationship with the tutor	Tutor may be reluctant to correct the student's knowledge deficits or reasoning skills
Be conventionally indirect (give hearer options and minimize disaggreements)	Student is not continually directed by the tutor	Students may not understand what is expected from them

Table 8

Speech Act Categories.

Speech Act	Definition	Example
Question (Q)	Interrogative or information seeking expression that is not an indirect request.	How does this piece fit here?
Reply to Question (RQ)	Response that specifically answers a previous question.	Yes, I'm fine.
Directive (D)	Request signaled by imperative form.	Give me the blue piece.
Indirect Directive (ID)	Request in non-imperative form.	Can you open the door?
Assertion (A)	Report about some state of affairs that could be true or false.	The risograph machine is broken.
Evaluation (E)	An expression of sentiment.	That stinks!
Verbal Response (R)	Spoken acknowledgement of the previous speech act.	Uh-huh.
Nonverbal Response (N)	Unspoken acknowledgement of the previous speech act.	(Head nod)
Juncture (J)	Lengthy pause in conversation.	(Pause)

Table 9

Performance of Four Connectionist Models of Speech Act Prediction.

	CORPUS		
	Children's Dyads	College Tutoring	Telephone Conversations
Maximum Activation Analysis			
Goodness-of-prediction Score			
Recurrent connectionist network	.289	.264	.358
Double-entry back propagation network	.292	.330	.330
Single-entry back propagation network	.268	.311	.291
Perceptron	.268	.331	.291
Hit rate (recurrent network)	.379	.451	.472
Base rate (recurrent network)	.122	.136	.178
Number of speech acts predicted	1	1	1
Above Threshold Analysis			
Goodness-of-prediction Score			
Recurrent connectionist network	.376	.420	.520
Double-entry back propagation network	.367	.420	.540
Single-entry back propagation network	.322	.364	.292
Perceptron	.320	.371	.292
Hit rate (recurrent network)	.565	.560	.696
Base rate (recurrent network)	.309	.242	.366
Number of speech acts predicted	1.8	1.5	1.9

Figure 1: The 5-Step Dialog Frame.

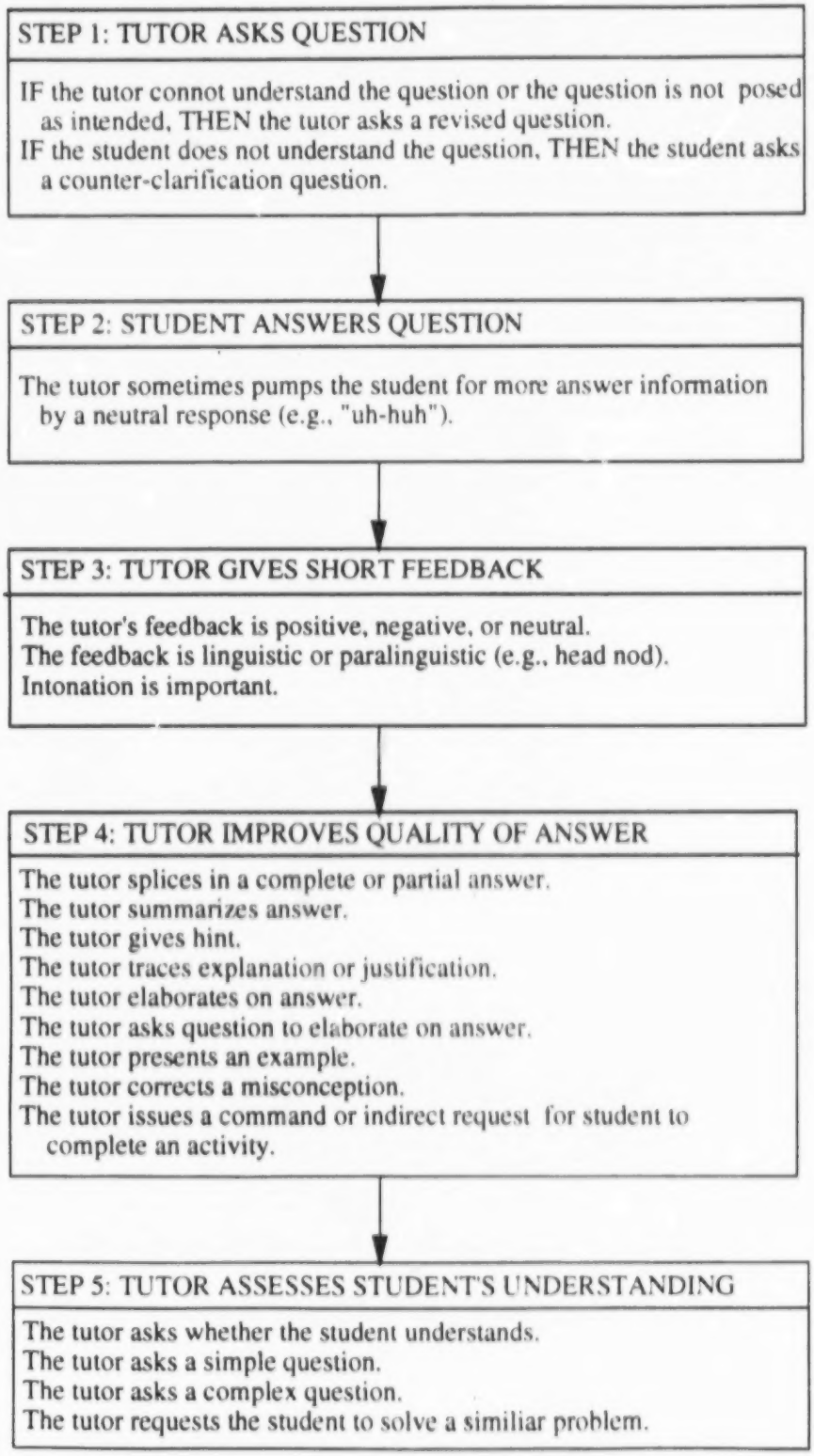

STEP 1: TUTOR ASKS QUESTION
IF the tutor connot understand the question or the question is not posed as intended, THEN the tutor asks a revised question.
IF the student does not understand the question, THEN the student asks a counter-clarification question.
STEP 2: STUDENT ANSWERS QUESTION
The tutor sometimes pumps the student for more answer information by a neutral response (e.g., "uh-huh").
STEP 3: TUTOR GIVES SHORT FEEDBACK
The tutor's feedback is positive, negative, or neutral.
The feedback is linguistic or paralinguistic (e.g., head nod).
Intonation is important.
STEP 4: TUTOR IMPROVES QUALITY OF ANSWER
The tutor splices in a complete or partial answer.
The tutor summarizes answer.
The tutor gives hint.
The tutor traces explanation or justification.
The tutor elaborates on answer.
The tutor asks question to elaborate on answer.
The tutor presents an example.
The tutor corrects a misconception.
The tutor issues a command or indirect request for student to complete an activity.
STEP 5: TUTOR ASSESSES STUDENT'S UNDERSTANDING
The tutor asks whether the student understands.
The tutor asks a simple question.
The tutor asks a complex question.
The tutor requests the student to solve a similiar problem.

Figure 2. Elman's Recurrent Network.

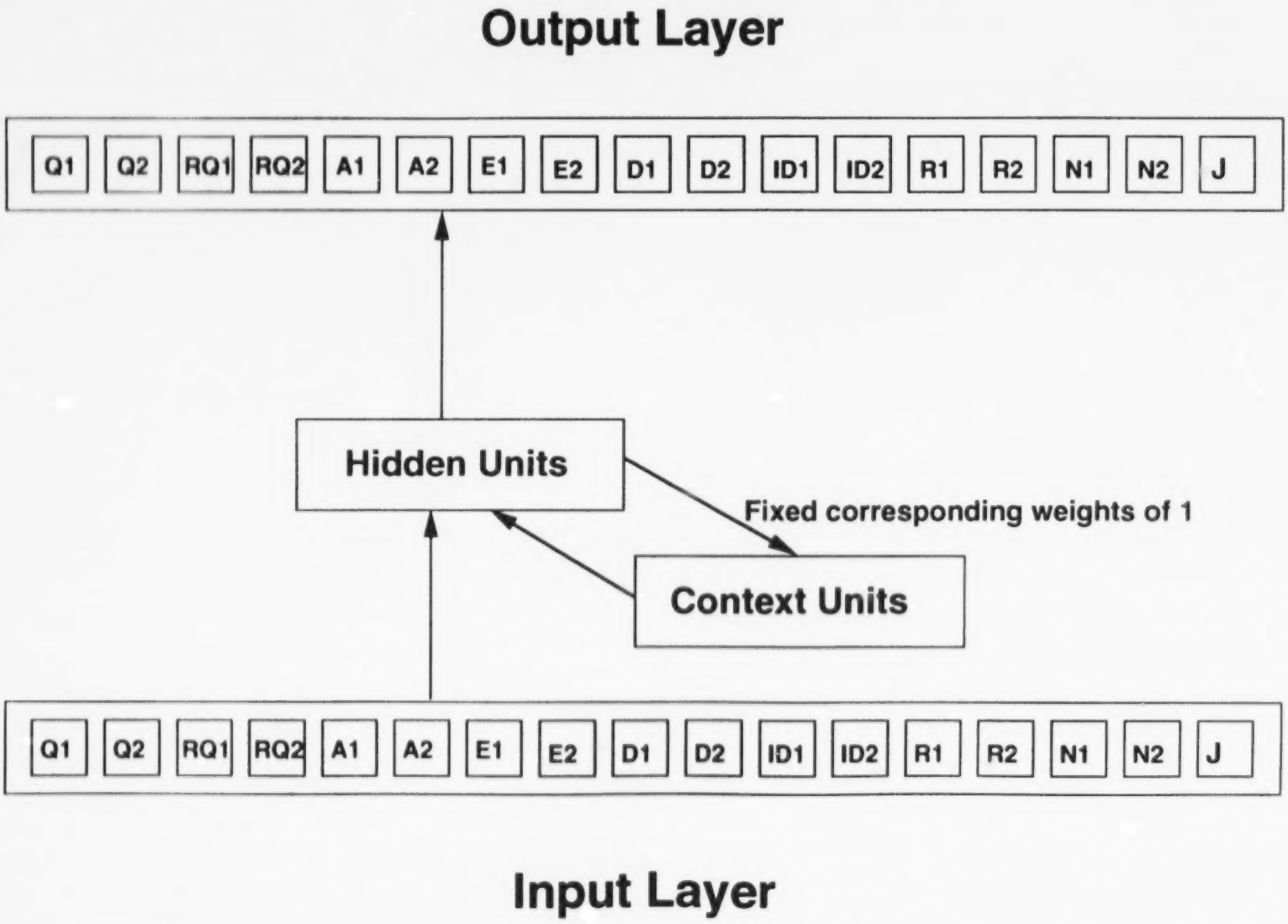
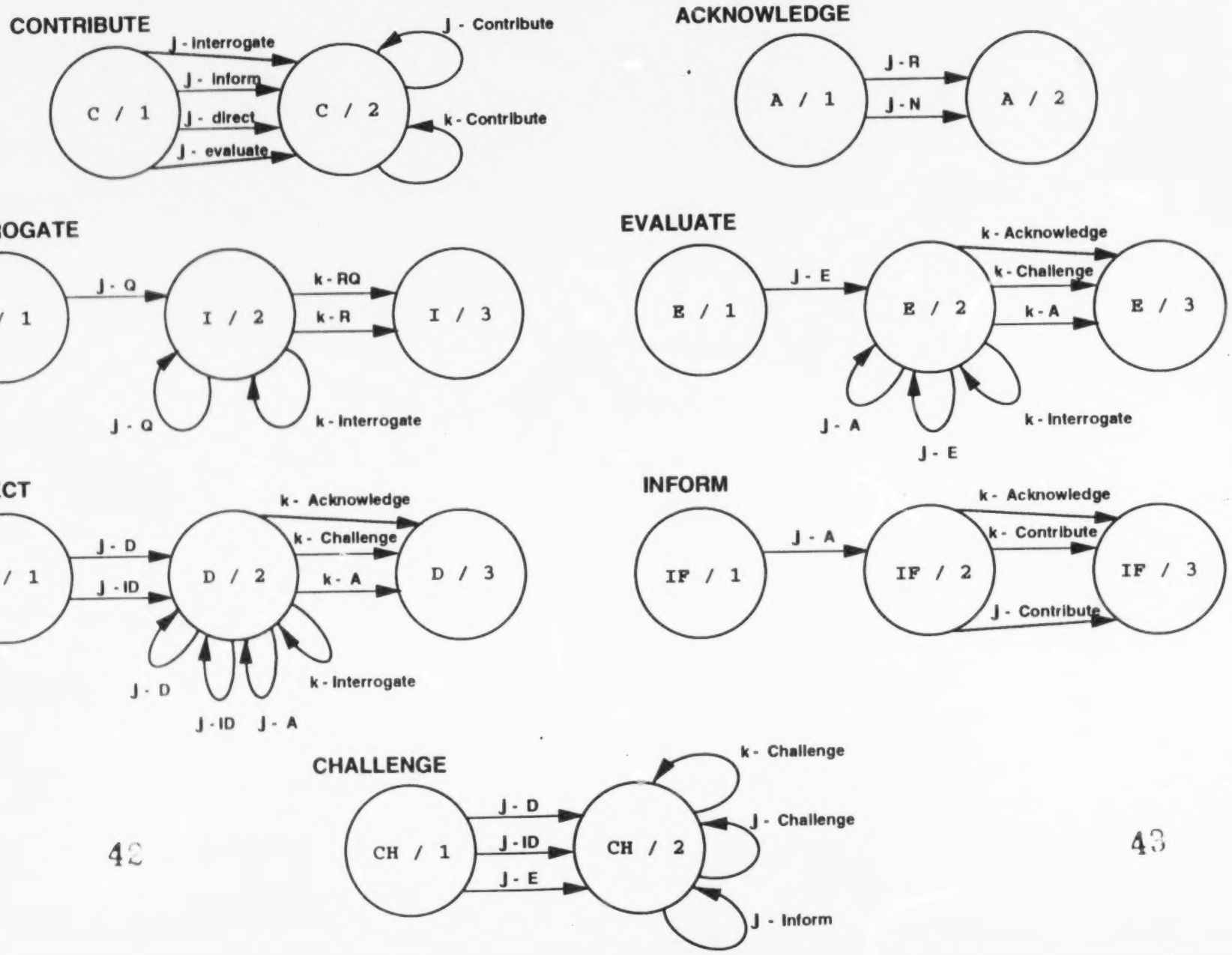


Table 3. Recursive Transition Network.



PUBLICATIONS AND PRESENTATIONS DURING PERIOD OF GRANT
(Graesser principle investigator)

Publications

Graesser, A. C. (in press). Knowledge representations and cognitive procedures. In G. Strube & K.F. Wender (Eds.), The cognitive psychology of knowledge: The German Wissenspsychologie Project (Advances in Psychology Series). Amsterdam: Elsevier (North Holland).

Graesser, A. C. (1993). Dialogue patterns and feedback mechanisms during naturalistic tutoring. Proceedings of the 15th Annual Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Graesser, A. C. (1993). Narrative thought and narrative language. Language and Speech, 36, 109-112.

Graesser, A. C., Byrne, P. J., & Behrens, M. L. (1992). Answering questions about information in structured databases. In T. Lauer, E. Peacock, & A. C. Graesser (Eds.), Questions and information systems. Hillsdale, NJ: Erlbaum.

Graesser, A. C., Gordon, S. E., & Brainerd, L. E. (1992). QUEST: A model of question answering. Computers and Mathematics with Applications, 23, 733-745.

Graesser, A. C., & Kreuz, R. J. (1993). A theory of inference generation during text comprehension. Discourse Processes, 16, 146-160.

Graesser, A. C., Langston, M. C., & Baggett, W. B. (in press). Exploring information about concepts by asking questions. In G.V. Nakamura, R.M. Taraban, & D. Medin (Eds.), Acquisition, representation, and processing of categories and concepts. Orlando, FL: Academic Press.

Graesser, A. C., Langston, M. C., & Lang, K. L. (1992). Designing educational software around questioning. Journal of Artificial Intelligence in Education, 3, 235-241.

Graesser, A. C., Magliano, J. P., & Haberlandt, K. (in press). Psychological studies of naturalistic text. In R. A. Zwaan & H. van Oostendorp (Eds.), Naturalistic text comprehension. Norwood, NJ: Ablex.

Graesser, A. C., Magliano, J. P., & Tidwell, P. (1992). World knowledge, inferences, and questions. In R. Beach, J. L. Green, M. L. Kamil, & T. Shanaham (Eds.), Multidisciplinary perspectives on literacy research. Urbana, IL: National Council of Teachers of English.

Graesser, A. C., & Marks, W. (1993). Models that simulate driver performance with hand controls. In B. Peacock & W. Karwowski (Eds.), Automotive ergonomics: Human factors in the design and use of automobiles. London: Taylor & Francis.

Graesser, A. C., & McMahen, C. L. (1993). Anomalous information triggers questions when adults solve problems and comprehend stories. Journal of Educational Psychology, 85, 136-151.

Graesser, A. C., McMahen, C. L., & Johnston, B. K. (in press). Questions asking and answering. In M. Gernsbacher (Ed), Handbook of Psycholinguistics. San Diego, CA: Academic Press.

Graesser, A. C., Person, N. K., & Huber, J. (1992). Mechanisms that generate questions. In T. Lauer, E. Peacock, & A. C. Graesser (Eds.), Questions and information systems. Hillsdale, NJ: Erlbaum.

Graesser, A. C., Person, N. K., & Huber, J. (1993). Question asking during tutoring and in the design of educational software. In M. Rabinowitz (Ed.), Cognitive science foundations of instruction (pp. 149-172). Hillsdale, NJ: Erlbaum.

Graesser, A. C., Person, N., & Johnston, G. S. (in press). Three obstacles in empirical research on aesthetic and literary comprehension. In R.J. Kreuz & M.S. MacNealy (Eds.), Empirical approaches to literature and aesthetics. Norwood, NJ: Ablex.

Graesser, A. C., Swamer, S. S., Baggett, W. B., & Sell, M. A. (in press). New models of deep comprehension. In B. K. Britton & A. C. Graesser (Eds.), Models of Understanding Text. Hillsdale, NJ: Erlbaum.

Kreuz, R. J. & Graesser, A. C. (1993). The assumptions behind questions in letters to advice columnists. Text, 13, 65-89.

Lang, K. L., Graesser, A. C., Dumais, S. T., & Kilman, D. (1992). Question asking in human-computer interfaces. In T. Lauer, E. Peacock, & A. C. Graesser (Eds.), Questions and information systems. Hillsdale, NJ: Erlbaum.

Langston, M. C., & Graesser, A. C. (1992). Question asking during learning with a point and query interface. Proceedings of the 14th Annual Cognitive Science Society, (pp. 921-926). Hillsdale, NJ: Erlbaum.

Langston, M. C., & Graesser, A. C. (1992). Student's questioning patterns as a function of time and goals. Proceedings of the Sixth National Conference on Undergraduate Research. Asheville, NC: University of North Carolina Press.

Lauer, T., & Graesser, A. C. (1992). Introduction. In T. Lauer, E. Peacock, & A. C. Graesser (Eds.), Questions and information systems. Hillsdale, NJ: Erlbaum.

Lauer, T., Peacock, E., & Graesser, A. C. (1992) (Eds.). Questions and information systems. Hillsdale, NJ: Erlbaum.

Long, D. L., Golding, J. M. & Graesser, A. C. (1992). Test on the on-line status of goal-related inferences. Journal of Memory and Language, 31, 634-647.

Magliano, J. P., Baggett, W. B., & Graesser, A. C. (in press). A taxonomy of inference categories that may be generated during the comprehension of literary texts. In R.J. Kreuz & M.S. Macnealy (Eds.), Empirical Approaches to literature and aesthetics. Norwood, NJ: Ablex.

Magliano, J. P., Bagget, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. Discourse Processes, 16, 35-53.

Magliano, J. P., Graesser, A. C., Eymard, L. A., Haberlandt, K., & Gholson, B. (1993). The locus of interpretative and inference processes during text comprehension: A comparison of gaze durations and word reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 704-709.

Magliano, J. P., Little, L. D., & Graesser, A. C. (in press). The impact of comprehension instruction on the calibration of comprehension. Journal of Reading and Instruction.

Millis, K., Graesser, A. C., & Haberlandt, K. (in press). The impact of connectives on memory for expository texts. Applied Cognitive Psychology.

Quinn, R. A., Houts, A. C., & Graesser, A.C. (in press). Moral questions and amoral answers: Who decides? Journal of Personality.

Quinn, R. A., Houts, A. C., & Graesser, A. C. (in press). Naturalistic conceptions of morality: A question answering approach. Journal of personality.

Quinn, R. A., Houts, A. C., & Graesser, A. C. (in press). Thinking about morality and moral thinking: Is there a distinction? Journal of Personality.

Swamer, S. S.,Graesser, A. C., Franklin, S. P., Sell, M. A., Cohen, R., & Baggett, W. B. (1993). A connectionist model of speech act prediction. Proceedings of the 15th Annual Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Presentations

Bertus, E. L., Baggett, W. B., Magliano, J. P., & Graesser, A. C. (June, 1992). The time course of inference generation and implications of consolidation. Presented at the American Psychological Society, San Diego, CA.

Graesser, A. C. (June, 1992). An explanation-based theory of inference generation during narrative comprehension. Symposium presented at the American Psychological Society, San Diego, CA.

Graesser, A. C. (June, 1992). Symposium on directions on discourse and text. Symposium presented at the Society for Text and Discourse, San Diego, CA.

Graesser, A. C., Kreuz, R. J., & Person, N. K. (June, 1992). The assumptions behind questions in discourse. Presented at the Society for Text and Discourse, San Diego, CA.

Langston, M. C. & Graesser, A. C. (June, 1992). Students' questioning patterns as a function of time and goals. Presented at the American Psychological Society, San Diego, CA.

Person, N. K. & Graesser, A. C. (June, 1992). Question asking in tutoring sessions. Presented at the American Psychological Society, San Diego, CA.

Person, N. K., Magliano, J. P. & Graesser, A. C. (June, 1992). The impact of reader goals, aesthetic experience, and literary aptitude on reading times for short stories. Presented at the Society for Text and Discourse, San Diego, CA.

Langston, M. C. & Graesser, A. C. (July, 1992). Question asking during learning with a point and query interface. Presented at the Fourteenth Annual Conference of the Cognitive Science Society, University of Indiana, Bloomington, IN.

Graesser, A. C. (November, 1992). Questioning mechanisms during tutoring. Colloquium presented at Southern Illinois University, Springfield, IL.

Magliano, J. P., Baggett W. B., & Graesser, A. C. (November, 1992). How long does it take to activate an inference. Presented at the Psychonomic Society Conference, St. Louis.

Graesser, A. C. (March, 1993). Mechanisms of question answering and question asking. Colloquium presented at the National Center for Health Statistics, Washington, D.C.

Graesser, A. C. (March, 1993). Tutorial interaction. Program report on Office of Naval Research grant, Princeton, NJ.

Graesser, A. C. (March, 1993). Commentary on symposium on "conceptual change". Meeting of the Society for Research in Child Development.

Baggett, W., Graesser, A. C., Franklin, S., & Swamer, S. (June, 1993). Computational models of speech act prediction. Presented at the Society for Text and Discourse, Boulder, CO.

Bertus, E., Baggett, W., Magliano, J., & Graesser, A. C. (June, 1993). Lexical versus unique situation-based inferences. Presented at the Society for Text and Discourse, Boulder, CO.

Graesser, A. C. (1993). Dialogue patterns and feedback mechanisms during naturalistic tutoring. Presented at the 15th Annual Cognitive Science Society, Boulder, Co.

Graesser, A. C., Magliano J. P., Person, N. K., & Kreuz, R. J. (June, 1993). Dialog patterns and questioning during naturalistic tutoring. Presented at the Society for Text and Discourse, Boulder, CO.

Langston, M., & Graesser, A. C. (June, 1993). Questioning patterns during knowledge exploration and learning. Presented at the Society for Text and Discourse, Boulder, CO.

Magliano, J., Johnston, S., Baggett, W., & Graesser, A. C. (June, 1993). Strategic inference generation. Presented at the Society for Text and Discourse, Boulder, CO.

Person, N. K., & Graesser, A. C. (June, 1993). Determining student's understanding in one-to-one tutoring sessions. Presented at the Society for Text and Discourse, Boulder, CO.

Swamer, S., Graesser, A. C., & Baggett, W. (June, 1993). Speech act prediction using a connectionist model. Presented at the Society for Text and Discourse, Boulder, CO.

Swamer, S. S.,Graesser, A. C., Franklin, S. P., Sell, M. A., Cohen, R., & Baggett, W. B. (1993). A connectionist model of speech act prediction. Presented at the 15th Annual Cognitive Science Society, Boulder, Co.

Zwaan, R., Dijkstra, K., & Graesser, A. C. (June, 1993). Comprehending literary narrative. Presented at the Society for Text and Discourse, Boulder, CO.

PUBLICATIONS OF CO-INVESTIGATOR

(1) STAN FRANKLIN

Franklin S. P., & Garzon M. (in press). Compatability via discrete neural nets. In D. E. Rumelhart, R. Smolenskey, & Moser (Eds), Mathematical perspectives on neural networks. Hillsdale, NJ: Erlbaum.

Franklin, S. P., & Garzon, M. (1992). On stability and solvability. Minds and Machines, 2, 71-83.

Garzon, M., Franklin, S. P., Baggett, W., Boyd, W., & Dickerson, D. (1992). Design and texting of general purpose neurocomputers. Journal of Parallel and Distributed Computing, 14, 203-220.

Swamer, S. S.,Graesser, A. C., Franklin, S. P., Sell, M. A., Cohen, R., & Baggett, W. B. (1993). A connectionist model of speech act prediction. Proceedings of the 15th Annual Cognitive Science Society. Hillsdale, NJ: Erlbaum.

(2) ROGER KREUZ

Graesser, A. C., & Kreuz, R. J. (in press). A theory of inference generation during text comprehension. Discourse Processes, 16, 146-160.

Kreuz, R. J., & Graesser, A. C. (1993). The assumptions behind questions in letters to advice columnists. Text, 13, 65-89.

Kreuz, R. J., & MacNealy, M. S. (Eds.). (in press). Empirical approaches to literature and aesthetics. In R. O. Freedle (series editor), Advances in discourse processes. Norwood, NJ: Ablex.

Kreuz, R. J., & Roberts, R. M. (in press). On satire and parody: The importance of being ironic. Metaphor and Symbolic Activity.

Kreuz, R. J., & Roberts, R. M. (in press). The empirical study of figurative language in literature. Poetics.

Kreuz, R. J., & Roberts, R. M. (1993). When collaboration fails: Consequences of pragmatic errors in conversation. Journal of Pragmatics, 19, 239-252.

Kreuz, R. J., Roberts, R. M., Johnson, B. K., & Bertus, E. L. (in press). Figurative language occurrence and co-occurrence in contemporary literature. In R. J. Kreuz & M. S. MacNealy (Eds.), Empirical approaches to literature and aesthetics. Norwood, NJ: Ablex.

Roberts, R. M., & Kreuz, R. J. (in press). Nonstandard discourse and its coherence. Discourse Processes.

Roberts, R. M., Kreuz, R. J., Gilbert, D. K., & Bainbridge, E. A. (in press). Discourse expectation and perceived coherence. In H. Van Oostendorp & R. A. Zwaan (Eds.), Naturalistic text comprehension. Norwood, NJ: Ablex.